

Discussion Paper No.331

海外における公的統計マイクロデータと行政記録情報の  
利活用の動向について

中央大学経済学部  
伊藤 伸介

August 2020



INSTITUTE OF ECONOMIC RESEARCH  
Chuo University  
Tokyo, Japan

# 海外における公的統計マイクロデータと行政記録情報の 利活用の動向について\*

中央大学経済学部  
伊藤 伸介

## 1. はじめに

一般に、公的統計データや行政記録情報は、統計表(集計結果表)あるいはマイクロデータ(個票データ)という形態で利用することが可能である。公的統計のマイクロデータについては、各国によって異なる様相を呈しているものの、利用目的、利用対象、利用場所、および利用の仕方によって、様々な形態でマイクロデータの作成・提供が行われている。それは主として、①学術研究目的のための利用者を対象として提供される個票データ(非識別データ、deidentified data)(例えば、Eurostat の Secure Use File 等)の提供サービス(リモートアクセス(remote access)も含む)、②リモートエグゼキューション(remote execution、オンデマンドによる集計サービス(リモート集計)も含む)、③主として学術研究のために作成される匿名化マイクロデータ(anonymized microdata)(例えば、イギリスのライセンス(End User

---

\* 本稿は、2018 年度統計関連学会連合大会(2018 年 9 月 於 中央大学)、2019 年度統計関連学会連合大会(2019 年 9 月 12 日、於 滋賀大学)における報告内容に基づいている。本稿の作成にあたって、アメリカセンサス局(2017 年 8 月 24 日)、カナダ統計局(2017 年 8 月 28 日)、ヨーロッパ統計局(Eurostat)(2019 年 3 月 11 日、木村有紀氏(NTT データ経営研究所)とインタビューを実施)、ノルウェー統計局(2019 年 8 月 14 日)、ノルウェーリサーチデータセンター(Norwegian Centre for Research Data=NSD)(2019 年 8 月 15 日~16 日)、ドイツ連邦統計局(2019 年 3 月 13 日)、イギリス国家統計局(2019 年 3 月 19 日および 2020 年 2 月 10 日。なお、2020 年 2 月 10 日については川崎のぞみ氏(NTT データ経営研究所)とインタビューを実施)においてそれぞれヒアリング調査を実施した。また、本稿の一部は、Richard Welpton 氏(Health Foundation)と Louise Corti 氏(Essex University, U. K. Data Service)へのインタビュー(2020 年 2 月 12 日と 13 日にそれぞれ実施)も含まれる。インタビューに応じていただいた、Tom Mule 氏(アメリカセンサス局)、Scott Konicki 氏(アメリカセンサス局)、Patrick Mason 氏(カナダ統計局)、Philippe Gagné 氏(カナダ統計局)、Crystal Sowards 氏(カナダ統計局)、Richard Trudeau (カナダ統計局)、Danielle Lessard 氏 (カナダ統計局)、Jean Pignal 氏(カナダ統計局)、Aleksandra Bujnowska 氏(Eurostat)、Wim Kloek 氏(Eurostat)、Maurace Brandt 氏(ドイツ連邦統計局リサーチデータセンター数理統計方法部門)と関係職員 3 名、Johan Heldal 氏(ノルウェー統計局)、Håkon Karlsen 氏(ノルウェー統計局)、Johan Sjøberg 氏(ノルウェー統計局)、Øyvind Langsrud 氏(ノルウェー統計局)、Diana-Cristina Iancu 氏(ノルウェー統計局)、Marianne Aamodt 氏(ノルウェー統計局)、Vigdis Kvalheim 氏(NSD)、Ørnulf Risnes 氏(NSD)、Knut Skjåk 氏(NSD)、Peter Stokes 氏(イギリス国家統計局)、Iain Dove 氏(イギリス国家統計局)、Richard Welpton 氏(Health Foundation)と Louise Corti 氏(エセックス大学)にお礼を申し上げたい。

Licence)を必要とするタイプのマイクロデータ(以下、「ライセンス型マイクロデータ」と呼称)等の作成・提供、および一般公開型マイクロデータ(public use microdata)(例えば、アメリカ人口センサスの PUMS(=Public Use Microdata Sample)やカナダの人口センサスにおける PUMFs(=Public Use Microdata Files)等)の公開という形で展開されている(伊藤(2016a, 2018b))。

図1は、マイクロデータの提供形態を具体的に整理したものである。第1の個票データは、氏名や住所といった直接的な識別子を含む原データ(raw data)から直接的な識別子を除くことによって作成される粒度の細かなマイクロデータである。このようなタイプのマイクロデータは、オンサイト施設やリモートアクセス(remote access)といった制御されたセキュアな環境によってアクセスすることができる。第2のリモートエグゼキューションは、利用者がリモートで作成・送付したプログラムを実行するか、あるいはオンデマンドの集計システム上で集計表に含まれる変数が選択されることによって、分析結果を出力する方法である。第3の匿名化マイクロデータは、個票データに対して各種の匿名化技法を適用することによって作成されるデータであって、例えば、ヨーロッパ諸国で作成・提供されている Scientific Use File(学術研究用ファイル、SUF)が該当する(伊藤(2018a))。第4の一般公開型マイクロデータは、個票データに対して各種の匿名化技法の適用によって匿名化の強度を高め、個体が特定されるリスクをできるだけ小さくすることによって作成されるマイクロデータである。一般公開型マイクロデータは、一般に利用目的に制限がなく、公表されている集計表(オープンデータ)と同様に、誰でも web からアクセス可能な形で、マイクロデータが提供される。その一方で、ヨーロッパ統計局(Eurostat)は、Public Use File(一般公開用ファイル、PUF)を作成・提供しているが、Eurostat の場合、利用上の登録を行った者のみが、PUF をダウンロードして利用することができる。また、Eurostat においては、SUF を利用する前のプログラム作成のためのテストデータや統計分析の利用に関する教育用のデータとしての利用が推奨されている(伊藤(2018a, p.85))。

わが国においても、現行の統計法制度の下で、統計調査の調査票情報(個票データ)の提供と匿名データの作成・提供が行われている。調査票情報については、統計法 33 条に基づいて、①磁気媒体による提供、②オンサイト施設の利用という2つの形態が存在する。しかしながら、わが国では、法制度的・統計技術的な観点から、リモートアクセスによる公的統計の個票データの利用サービス、リモートエグゼキューションによるプログラム送付に基づく分析結果の提供、オンデマンド集計システムによる集計結果の提供については、これまで実用化に向けた十分な検討はなされてきたとは言えない。こうした状況を考慮した場合、海外の公的統計マイクロデータの作成・提供の現状を把握することによって、今後のわが国のマイクロデータの提供の展開方向を模索するための参考にすることが可能になる。

さらに、わが国においては、行政記録情報の二次利用に対する社会的な関心が高まっていることから、わが国でも、法制度な措置と秘密保護に関する統計技術的な手法の両面から行政記録情報の利用可能性を検討することが求められる状況にある。これに関しては、行政記録情報に基づく統計の作成が展開されている北欧諸国(フィンランド、スウェーデン、ノルウェー、デンマーク等)を中心に、海外における行政記録情報の二次利用の現状を把握する必要があると考える。

本稿では、最初に、諸外国における公的統計マイクロデータの作成・提供の現状を明らかに

する。つぎに、主としてヨーロッパ諸国と北米諸国を対象に行政記録情報の利活用の状況について概括する。最後に、これらの海外における公的統計や行政記録情報の利用の現状を踏まえた上で、わが国における今後の展開方向について述べることにしたい。

## 2. 諸外国における公的統計マイクロデータの作成・提供の現状 —Eurostat、ドイツとイギリスの例—

本節では、諸外国における公的統計のマイクロデータの作成・提供状況について論じることとする。表 1 は、主要な欧米諸国の統計作成部局が行っているマイクロデータの提供状況を示したものである。個票データ(非識別データ)は、一部の国を除けば、欧米各国でオンサイト施設のようなセキュアな環境で提供されている。イギリス、フランス、オランダ等のヨーロッパの多くの国々では、個票データのリモートアクセスが行われている。プログラム送付型のリモートエグゼキューションによる提供が行われているのは、ドイツ、カナダ、ノルウェーであって、リモートエグゼキューションによる提供システムは存在するものの、実質的な運用がなされていない国(例えば、オランダ等)もある。さらに、オンデマンドの集計システムによって分析結果が提供される国々として、オランダ、カナダ、オーストラリアなどを指摘することができる。

表 1 では、個票データから匿名化処理が施されたデータが、匿名化の強度によって、①匿名化マイクロデータと②一般公開型マイクロデータに類別される。匿名化マイクロデータはヨーロッパの多くの国々で作成・提供されている。それに対して、一般公開型マイクロデータが広範に作成・提供されているのはアメリカとカナダである。ヨーロッパ諸国は、主に教育目的やテストデータの利用を想定した上で、一部の統計調査が PUF として提供されている。

以下では、Eurostat、ドイツとイギリスにおける公的統計マイクロデータの提供の現状について概括する。

### 2-1. Eurostat におけるマイクロデータの提供について

Eurostat では、現在一般公開型ファイル(Public Use Files=PUFs)、学術研究用ファイル(Scientific Use Files=SUFs)および Secure Use File が作成・提供されている。SUFs については 2002 年に提供が開始された。European Union Labour Force Survey(EU-LFS)等の 12 調査が CD-ROM の形で提供されているが、Eurostat の web サイトからのダウンロードが可能な状況である。12 調査の中に、Structure of Earnings Survey 等の事業所・企業系の SUF が存在する。また、Secure Use File については、Community Innovation Survey といった事業所・企業系の個票データのみが Eurostat のオンサイト施設(Safe Centre)でのみアクセスすることが可能である。なお、Eurostat が提供している SUFs や Secure Use File の申請を行うためには、(EU 域外の研究者も含む)研究者の所属機関が、「承認された組織(recognized organization)」であることが求められる。この「承認された組織」という概念は、ヨーロッパ諸国の多くで採用されている、個票データの提供サービスを展開するための要件の 1 つとなっている(伊藤(2018))。

また、Eurostat では、2016 年から EU-LFS と EU-SILC(=Statistics on Income and Living Conditions)の PUFs が公開されており、Eurostat の CROS(=Collaboration in

Research and Methodology for Official Statistics)と呼ばれるポータルサイトから PUFs の取得が可能になっている。PUFs については利用目的に対する制限は当てはまらないが、Eurostat の場合、PUFs の取得において、①Eurostat のポータルサイトにおいて利用者の登録が求められていること、②教育目的としての利用やプログラム作成のためのテストデータとしての利用が推奨されている。

Eurostat におけるマイクロデータの提供サービスにおいて、利用者からの SUFs の評判はよくなかったことが知られている。また、オンサイト利用後の分析結果のチェックは Eurostat にとって負担となっていた。さらに、分析結果のチェックの負担が大きく、人的リソースがかかり、システム開発やそのメンテナンスに係る費用も高いことを理由に、Eurostat はリモートアクセスによる個票データの提供サービスも行っていない。こうした状況が、Eurostat において Confidentiality on the fly と呼ばれるリモートエグゼキューションのシステムの導入に向かわせる契機となった。

Confidentiality on the fly の構想において参考にしたシステムは、オーストラリア統計局が開発されたオンデマンド集計システムである TableBuilder(伊藤・谷道・小島(2018))である。TableBuilder によって提供されるサービスは、オンデマンドの集計表の作成のみであって、回帰分析のサービスは行われていないが(2020年4月時点)、Confidentiality on the fly も同様にオンデマンドの集計表の作成・提供を想定している。Confidentiality on the fly での利用の対象となるマイクロデータは、世帯・人口系のデータのみであって、EC-LFS と EU-SILC の 2 つが計画されている(2019年3月時点)。事業所・企業系のデータについては、攪乱的手法を適用しようとする、誤差が大きくなる懸念があるため、事業所・企業系のデータは、Confidentiality on the fly で利用可能なデータに含まれていない。また、Eurostat は Confidentiality on the fly において、数量表(magnitude table)ではなく、度数表(frequency table)の作成と提供を検討している。なお、Eurostat には他のヨーロッパ諸国から労働力調査等の個票データが移送されるが、このデータには直接的な識別子が削除されている。そのため、他の個票データをリンクするための ID が存在しないことから、リンクされた個票データの基づくオンデマンド集計は困難である。

## 2-2 ドイツ連邦統計局におけるマイクロデータ提供の現状

ドイツ連邦統計局におけるマイクロデータの提供形態は、①Scientific Use File(SUF)の提供、②Public Use File(PUF)の公開、③オンサイト施設による個票データの提供サービス、および④プログラム送付型のリモートエグゼキューションに大別される。

ドイツにおける SUF は、「事実上の匿名性(factual anonymity)」の概念に基づいて、攪乱的手法を含む秘匿処理を施すことによって作成された匿名化マイクロデータである。事実上の匿名性とは、「著しく大きな時間、経費および労力の支出によって、当事者に関連づけることができない」こと(濱砂(1999))であって、それは、現行の連邦統計法第 16 条「秘密保護」第 6 項にも「事実上の匿名化が施された個票データ(de facto anonymised individual data)」と定義されている。事実上の匿名性に基づく匿名化マイクロデータの作成にあたって、標本調査であるマイクロセンサス(Microcensus)を用いて、事実上の匿名性について実証的に明らかにするために、1980 年代に実施された外部情報とのマッチングによる実験の成果(例えば、Müller et al.(1995))も、定量的な参考情報になっていると言える。さらには、秘匿性

や有用性に関する実証的な研究もドイツ連邦統計局内部では行われてきた。

ドイツにおける PUF は、「申告義務者ないしは当事者に、もはや関係づけることができない」ように匿名化を施す絶対的な匿名性(absolute anonymity)の考え方に基づいている(濱砂(1999))。ドイツ連邦統計局は、秘匿処理の方法としてサンプリングとスワッピングを適用することによって、PUF を作成している。なお、他国の事例とは異なり、ドイツ連邦統計局では、PUF の提供は、250 ユーロと有料となっている。同様に、絶対的な匿名性の考え方をもとにして、PUF とは異なる教育用のデータである Campus File が提供されている。さらに、利用者がプログラムコードを作成するための data structure file も絶対的な匿名性に基づいて作成されている。

ドイツでは、連邦統計局のオンサイト施設や連邦州の Safe Center(16 か所)において、150 の統計調査に関する個票データの利用が可能になっている(2019 年 3 月時点)。これは、連邦統計法第 16 条「秘密保護」第 6 項に基づく。すなわち、連邦統計局のオンサイト施設や safe center のサーバにおいて安全な環境が確保されている場合に「形式的に匿名化された個票データ(formally anonymized individual data)」へのアクセスが可能になる。オンサイト施設における個票データのリンケージに関しては、事業所・企業系のデータについては ID によるリンケージが可能であるのに対して、世帯・人口系のデータ(ex. Microcensus 等)に関しては共通 ID がないため、他の調査とは統計的マッチングでのみリンケージが可能な状況である。さらに、オンサイト施設や Safe Center で個票データに基づく分析結果のチェックも行われる。

ドイツでは、例えば大学の研究室からリモートアクセスの形で個票データにアクセスすることはできない。それは、ドイツでは、個票データの移送が法律上認められていないからであって、リモートアクセスは、データの移送に該当するために、ドイツにおいては法的な観点からリモートアクセスによる個票データの提供サービスは行われていない。

こうしたことから、ドイツ連邦統計局は、2001 年からプログラム送付型のリモートエグゼキューションを実施してきた。また、ドイツ連邦雇用庁(IAB)でも、プログラム送付型のリモートエグゼキューションによる提供サービスが存在する。ドイツ連邦統計局では、SAS、SPSS、STATA および R のプログラムコードを送付することが可能になっている。なお、料金は 1 データあたり 250 ユーロであり、年間単位で使用するデータの数によって使用料が異なるように設定される。ただし、ドイツ連邦統計局によれば、プログラム送付型のリモートエグゼキューションの場合、分析に用いるプログラムや分析結果のチェックに時間を要することから、効率的な方法とは言えないとのことであり、ドイツ連邦統計局からはリモートエグゼキューションは推奨されなかった。

### 2-3 イギリスにおけるマイクロデータ提供の現状

イギリスでは、公的統計マイクロデータの作成・提供は、イギリス国家統計局(The Office for National Statistics=ONS)を中心に行われている。具体的には、①オンサイト施設による個票データの提供サービス、②リモートアクセスによる個票データのアクセス、③ライセンス型マイクロデータの提供、④教育用の公開型マイクロデータの公開という 4 つのチャンネルが存在する(伊藤(2012, 2014, 2016b))。これらのマイクロデータの提供状況の概略図を示したものが、図 2 である。なお、図 2 で示されている行政記録情報の利活用については次節で

述べることにしたい。

ONS のオンライン施設である Secure Research Service (SRS、以前は Virtual Microdata Laboratory と呼ばれていた) において、公的統計の個票データが利用可能である。イギリス統計法(The Statistics and Registration Service Act 2007)に基づき、承認された研究者(approved researcher)の資格を取得すれば、公的統計の個票データにアクセスすることができる。LS データ(ONS Longitudinal Study of England and Wales)だけでなく、リンクされた公的統計の個票データに関しても SRS においてアクセスすることができる。なお、個票データにおける分析結果については、研究者が外部に持ち出す前に、ONS の担当者によって、安全性のチェックが行われる。ONS の担当者が分析結果のチェックを行う場合、SRS の担当部局が、統計表やマイクロデータの秘匿に関する部局と協力しながら、マニュアルによるチェックを行っている。

「承認された研究者」の資格を取得するための申請書には、申請する研究プロジェクトに関して、①研究目的、②個票データを利用する期間、③個票データの利用経験、④申請者の所属機関、⑤研究成果について公表する論文等の記載が求められる(伊藤(2018a))。個票データの利用経験の有無に関わらず、研究プロジェクトで提示される研究成果が、公共財(public goods) になるかどうかを ONS 内で検討した上で、容認された場合に個票データのアクセスが可能になる。

さらに、イギリスでは、UKDS の the Secure Lab がリモートアクセス施設を持っており、研究者は自身の研究室からリモートで個票データ(非識別データ)のアクセスを行うことができる。事業所・企業を対象とした約 200 の公的統計等の個票データ(非識別データ)が、UK Data Service のサーバに保管されており、研究者がアクセスすることが可能になっている。

イギリスにおけるライセンス型マイクロデータについては、エセックス大学の U.K. データアーカイブ(U.K. Data Archive=UKDA)が保有する社会調査データだけでなく、公的統計の匿名化マイクロデータについても、UK Data Service(UKDS)において提供サービスが行われている。具体的には、イギリス国家統計局が実施する労働力調査(LFS)等のサーベイマイクロデータ、人口センサスの匿名化マイクロデータ(1991 年、2001 年における SARs(Samples of Anonymised Records))等が該当する。人口センサスについては、1991 年に匿名化マイクロデータの提供が開始されて以降、複数のタイプの匿名化マイクロデータが作成されてきた。2021 年の人口センサスにおいても、個人・世帯属性が詳細なマイクロデータと、地域情報が詳細なマイクロデータの 2 種類の匿名化マイクロデータの作成が計画されている。

海外では、人口センサスの匿名化マイクロデータを作成する場合に、攪乱的手法(perturbative method)であるスワッピング(data swapping)を用いることがあるが、イギリスでは、集計結果表(統計表)の元になる個票データに対してスワッピングが適用されていることが特徴的である。ONS は、2001 年人口センサスにおいてランダム・スワッピング(random swapping)を適用していた。それに対して、2011 年の人口センサスでは、スワッピングの方法をターゲット・スワッピングに変更している。ONS では、2021 年人口センサスにおいても、ターゲット・スワッピングが適用されることが計画されている。

ライセンス型マイクロデータを学術研究目的で利用する場合、研究者・学生は、UKDS において利用者登録を行い、End User License Agreement に同意すれば、無料でライセンス型マイクロデータをダウンロードすることが可能である。また、学術研究目的だけでなく、教

育目的で利用することも可能であるが、データに含まれる個人情報漏洩した場合、学生であっても罰則の対象となる。ライセンス型マイクロデータの中には、小地域分析用のマイクロデータも存在する(例えば、2001年の人口センサスの小地域マイクロデータ(Small Area Microdata)等)。その場合、基本的な地域の最小単位は行政官庁区域(government office region)である。さらに、ライセンス型マイクロデータに関しては、商用目的であっても、有料で利用可能な公的統計マイクロデータ(例えば、Annual Population Survey 等)が存在する。

ONSは、2011年人口センサスにおいて、教育用の公開型マイクロデータ(Public Use File)の作成・提供を開始した。教育用の公開型マイクロデータの作成にあたっては、サンプリングやスワッピングを使用していることが知られている。このマイクロデータは、ONSのウェブページでダウンロードをすることが可能である。なお、2021年の人口センサスでも、教育用の公開型マイクロデータの作成が計画されている。

### 3. 海外における行政記録情報の利活用の現状—ヨーロッパと北米諸国の事例—

北欧諸国(フィンランド、スウェーデン、ノルウェー、デンマーク)、さらにはオランダといった国々では、行政記録情報に基づいた統計作成システムが確立されてきた。これらの国々は、出生あるいは移住した時点で、個人に識別子としてのIDが付与されており、このIDを介して、各種の行政記録情報のリンケージが可能になっている。それによって、基本的な人口社会情報だけでなく、所得、社会補償給付、教育、雇用、医療・健康等に関する情報が行政記録情報として把握することが可能になっている。

デンマークやノルウェーといった北欧諸国においては、様々なレジスターが統計局に自動で集められており、社会保障番号(social security number)が仮名化された形で統計局に保管されている。レジスターベースの統計システムにおいては、行政記録情報が公的統計作成のための原データであって、この原データから直接的な識別子を除いた非識別データ(個別データ、deidentified data)がマイクロデータとして提供されている。学術目的のための非識別データの広範な利活用が可能になっている。

それに対して、レジスターベースではなく、統計調査を実施することによって公的統計を作成しているイギリスでは、個人のIDが存在しない。しかしながら、イギリスにおいては、名前や住所といった直接的な識別子を用いて仮名化されたIDの作成と、仮名化されたIDを用いたリンケージの方法論に関する研究が進められている。また、イギリス国家統計局において、行政記録情報のリンケージに基づいた人口センサスデータの作成に関するプロジェクトが展開されてきた。

本節では、ヨーロッパと北米諸国を対象に行政記録情報の利活用の現状について見ていくことにしたい。具体的には、デンマークとノルウェーとイギリスにおける行政記録情報の利用状況を述べる。さらにイギリスと同様に調査票ベースの統計作成を行っているアメリカとカナダを対象に、北米諸国における行政記録情報の利活用の状況を明らかにする。

#### 3-1 デンマークにおける行政記録情報の二次利用の動向について

デンマークは、個人に関する属性情報が行政記録によって把握されており、デンマークに



居住する個人に付与された個人識別番号によって、レジスターベースで記録されている様々な個人情報のリンケージが可能になっている。これらのリンクされた個人に関する行政記録情報は仮名化され、個人識別番号が個人の特長ができない別の ID に置換された非識別データとして利用することができる(伊藤(2017))。

デンマークにおける行政記録情報の二次利用の大きな特徴は、リモートアクセスを通じてのみ行政記録情報の利用サービスを行っていることにある。リモートアクセスサービスは、2001年より開始されており、匿名化マイクロデータやオンデマンド集計システムは存在しない。認証されている所属機関の研究者およびその共同研究者であれば、認証機関から非識別データにアクセスすることができる。さらに、デンマークの場合、研究者の研究室の外からでも、ラップトップの PC でリモートアクセスが可能である。認証機関に所属する研究者が外国に滞在している場合でも、リモートアクセスを行うことができる(伊藤(2017))。なお、2019年2月現在、2000以上の研究プロジェクトによって、学術研究のための行政記録情報の利用が行われている。

行政記録情報の利用を希望する場合、利用申請の手続きを行った上で、誓約書に署名することによって、利用することができる。人口社会情報、所得情報、教育情報、労働市場や企業経済に関する情報だけでなく、医療・健康情報へのアクセスも可能である。そして、これらの様々な分野の情報が ID を通じてリンケージされたデータが学術研究目的で利用可能になっているのが、デンマークにおける北欧諸国における行政記録情報の二次利用の特徴だと言える。

デンマーク統計局で行われている秘匿処理は、直接的な識別子(氏名、住所)の削除と ID の仮名化のみである。したがって、属性情報に対して匿名化技法の適用はなされていない。デンマーク統計局と利用者との間の信頼関係がこうした仮名化のみの処理を可能にしているが、デンマーク統計局は、「知る必要性(need to know)」原則に基づいて、学術研究のために、利用者において必要な属性群のみを含む非識別データにアクセスすることを可能にする措置を取っている。また、2018年デンマークデータ保護法(Danish Act on Protection of Data 2018)、1985年行政機関法(Public Administration Act 1985)、デンマーク刑法(Danish Penal Code)は、行政記録情報の提供にあたって、個人情報の保護および個人情報の漏洩についての罰則に関する法的な根拠となっている。

デンマークで非常に興味深い点は、2018年にデンマーク統計局が、実証分析によって得られた成果物について研究者自身が自動的にチェックを行うことを可能にするスキャンシステムを開発したことである。これによって、研究者は、非識別データに含まれる個体が特定されないようにした上で、利用者の責任で分析結果を持ち出すことができる。分析結果のチェックの基準としては、主として、作成した集計表において、セルに含まれる度数は3以上であるといった閾値に関する基準がある。研究者がスキャンシステムで分析結果をチェックした後に持ち出す最終成果物については、デンマーク統計局のチェックがなくても論文や学会報告等でその成果物を載せることができる(伊藤・木村(2019))。それに対して、デンマーク統計局の担当職員が、すべての分析結果の中の一部をランダムに選別し、事後的に人手によるチェックを行っている。

デンマークにおけるこうした行政記録情報の提供状況に関しては、基本的には、デンマークの統計作成部局と研究者との間の信頼関係が重視されていると言えよう。

### 3-2 ノルウェーにおけるマイクロデータの提供状況について

ノルウェーにおいては、様々なレジスターに保管されている行政記録情報がノルウェー統計局に自動で集められており、社会保障番号 (social security number) を含む行政記録情報が仮名化された形で統計局に保管されている。仮名化された ID に基づいてリンケージが施された非識別データの作成が可能になる。この ID をさらに別の擬似 ID に変換した上で、ノルウェー統計局は、学術研究目的に対して利用者に行政記録情報を提供している。これは、ID に基づいて行政記録情報のリンケージが可能なデンマーク等の他の北欧諸国と同じ提供状況と言える。

ノルウェーにおけるマイクロデータに関しては、ノルウェー統計局が実施している、①行政記録情報の非識別データの提供と②プログラム送付型のリモートエグゼキューション、および③ノルウェーリサーチデータセンター (NSD)による統計調査の匿名化マイクロデータの提供の3つの形態が展開されている。

ノルウェーの場合、研究者が、研究プロジェクトに入っていれば、ノルウェー統計局内部のセキュアな環境で非識別データにアクセスするか、研究機関における local なサーバでデータを保管した上で、研究室で非識別データを利用することができる。ただし、利用申請から提供までに数か月の期間がかかるだけでなく、データの使用料がかかるとのことである。ノルウェーにおいても、「知る必要性」原則を踏まえて、マイクロデータの提供を行っている。その場合、どの変数を使うのかについて、事前にノルウェー統計局の担当者に相談する必要がある。また、ノルウェーの場合、リモートアクセスのみで行政記録情報の利用サービスが行われているデンマークとは異なり、リモートアクセスシステムは存在しておらず、オンサイト施設を備えたリサーチデータセンターも設置されていない。

ノルウェー統計局の興味深い動きとして、リモートエグゼキューションシステムである [microdata.no](http://microdata.no) を指摘することができる(Heldal et al.(2019))。ノルウェー統計局は2018年に Research Council Norway の資金援助を受けて、プログラム送付型のリモートエグゼキューションを開始した。2019年8月時点では、試行提供中であることから、[microdata.no](http://microdata.no) のリモートエグゼキューションのサービスは、無料で行われている。

プログラム送付型リモートエグゼキューションの開発が進められてきた背景としては、大学院生でも非識別データにアクセスできる仕組みを考案しようとしたことが指摘される。[microdata.no](http://microdata.no) は、2012年からノルウェー統計局と NSD との共同で進められてきた (RAIRD(=Remote Access Infrastructure for Register Data)プロジェクト)。プログラム送付型リモートエグゼキューションの方法論については、ノルウェー統計局が検討を進め、NSD が方法論に関する実証実験を行い、両者の緊密な関係において、リモートエグゼキューションの実装を行うことによって、[microdata.no](http://microdata.no) による提供サービスが実現した。

[microdata.no](http://microdata.no) にログインするためには、ノルウェーの国民番号が必要である。ログインすると、[microdata.no](http://microdata.no) では、124 個の変数と 1000 万のレコードを含むデータを用いて分析することができる。[microdata.no](http://microdata.no) のようなプログラム送付型のリモートエグゼキューションの特徴は、利用者が非識別データを閲覧することができないことにある。そこでプログラムを作成するためのメタデータが必要となる。[microdata.no](http://microdata.no) のメタデータは、NSD のようなデータアーカイブ施設において社会調査データに関して公開されるメタデータと変わら

ない情報量を備えている。また、各種のデータに含まれる変数名の変遷に関する情報もメタデータとして利用することが可能である。特徴的なのは、microdata.no で利用可能なデータに含まれる変数について分類区分の変遷がわかることである。さらに、基本統計量も把握可能である。

microdata.no では、プログラムを動かして得られた分析結果を rule based approach(伊藤(2016c))に基づきチェックの自動化が行われている。そのために、以下のような秘匿処理が取られている。

- ①分析されたサンプルサイズが最低 1000 以上である。
- ②度数表および数量表にいずれについても、すべてのセルに対してノイズが付与されている。これは、オーストラリア統計局の TableBuilder の影響が強いと言える。
- ③外れ値への対応として、すべての量的変数の上位 1%と下位 1%のデータを削除する。

図 3 は、ノルウェーにおけるマイクロデータの提供のタイプと有用性・秘匿性の関係を示したものである。データの有用性と個人情報の特定制のリスクだけでなく、個人情報の特定制に要する費用の観点から、マイクロデータが位置づけられる。非識別データが、特定制のリスクが高く、有用性も高いデータとなるのに対して、Public Use File は、特定制のリスクは低くなるが、有用性が低くなるデータとみなされる。これらに対して、microdata.no の提供システムについては、個人情報の特定制のリスクは Public Use File と同じレベルで低くなるが、有用性に関しても、非識別データのアクセスほどではないもののデータの有用性が確保されている。なお、このフレームワークは、先述の Five Ssfes Model に基づいていることに留意されたい。

ノルウェーのデータアーカイブ施設である NSD は、社会調査データの保存と共有(提供)を行っているが、公的統計についても、標本調査データを対象に匿名化マイクロデータの提供を行っている。なお、各種の公的統計の匿名化マイクロデータは、NSD の web 上にあるデータカタログを見てダウンロードすることが可能になっている。

社会調査データに関するメタデータの整備も NSD で行われている。また社会調査データや匿名化マイクロデータを検索するためのデータカタログも NSD のウェブサイトを利用可能になっている。なお、海外におけるメタデータの整備状況については伊藤・前田(2019)を参照されたい。

### 3-3 イギリスにおける行政記録情報の二次利用の現状

本節では、イギリスにおける学術研究のための行政記録情報の利活用の現状についても紹介することにした。イギリスでは、近年行政記録情報の二次利用が展開されており、行政記録情報と公的統計のマイクロデータとのリンケージがなされた行政記録情報や行政記録情報同士でリンケージされたデータの提供も行われている。当初は、Administrative Data Research Network (ADRN) というプロジェクトとして、エセックス大学の Administrative Data Service(ADS)を通じて、Administrative Research Data Centre(ADRC)において、行政記録情報が利用可能になっていた(伊藤(2016b))。ADRN プロジェクトは、2018 年に終了し、後継のプロジェクトとして、現在、ADR UK という新たなプロジェクトが展開されて

いる。このADR UKには、イングランド、ウェールズ、スコットランド、北アイルランドのそれぞれにおいて、行政記録のリンケージのサービスが展開されており、イングランドでは、ONSが一元的に行政記録のデータの利用申請の受付、行政記録情報を保有する政府機関との交渉、行政記録情報間のリンケージおよびリンクされた行政記録情報の提供を担っている。なお、イギリスでは、北欧諸国のような個人を識別する共通のIDが存在しないことから、名前、住所といった直接的な識別子、さらには性別や出生日等を使用して、行政記録情報間のリンケージが行われている。

イギリスにおける行政記録情報の利用は、2017年4月にDigital Economy Act 2017の法案が議会を通過し、2018年に法律が施行されたことによって新たに展開される(伊藤(2018a))。Digital Economy Act 2017においては、研究目的のために、各種の統計データ、行政記録情報、さらには政府によって(潜在的に)生み出されるビックデータの利用(リンケージされたデータの利用も含む)の可能性が指向されている。具体的には、「第5章 研究目的のための共有」中の「第64条 研究目的のための情報の開示(disclosure)第1項」において、「公的な当局(public authority)によって保有される、公的当局と関連している情報は、実施されている研究の目的のために、他の個人に開示される」ことが明記されている。リンケージされた行政記録情報の提供は、このDigital Economy Act 2017に基づいて行われている。なお、行政記録情報の秘密保護に関しては、一般データ保護規則(General Data Protection Regulation=GDPR)が、その法的な根拠となっている。

Digital Economy Act 2017が施行されてからも、行政記録情報の利用に関する基本的な考え方は変わらない。それは、Five Safes modelによるデータ提供の枠組み(Ritchie(2008), 伊藤(2016c))に基づいているからである。Five Safes modelは、①安全なプロジェクト(safe projects)、②安全な利用者(safe people)、③安全なデータ(safe data)、④安全な施設(safe settings)および⑤安全な分析結果(safe outputs)の5つの基準から構成される。

Five Safes modelの枠組みによれば、Digital Economy Act 2017においては、安全な利用者を表す accredited researcher(ONSによる「承認された研究者」と同様の概念)の資格を取得することが必要である。そのため、特定の研究プロジェクトが行政記録情報の利用申請を行うと、安全な研究プロジェクトかどうかについてUK Statistics AuthorityのPanelによる審査を受けることが求められる。公的統計の個票データと同様に、行政記録情報の利用によって公共財としての研究成果が得られることも、accredited researcher取得の要件と言える。さらに、accredited researcherの取得にあたっては、行政記録情報を利用するためのトレーニングが義務付けられている。そして、行政記録情報の利用の認可が得られると、ONSのSRSといったオンサイト施設、UK Data Serviceのリモートアクセス施設のような安全な施設においてのみ行政記録情報の利用が可能になる。行政記録情報を用いて作成された記述統計量、集計表や回帰分析といった分析結果のチェックについては、公的統計の個票データにおけるチェック基準と同様の基準が適用されており、集計表のセルの閾値といったルールに厳格に従う rule based approachではなく、提供者と利用者との相互の信頼関係に基づいて最適な分析結果の持ち出しを可能にする principle based approachが採用されている。

### 3-4 アメリカセンサス局における行政記録情報の利用状況

本節においては、アメリカセンサス局における行政記録情報の活用状況を概観する。アメリカセンサス局は、MAF(=Master address file)と呼ばれる住所情報とTIGER(=Topologically Integrated Geographic Encoding and Referencing)と呼ばれる地理的情報を有している。MAFとTIGERは、アメリカセンサス局内部において連結されており、MAF/TIGERとして一体化されている(森(2007))。具体的には、それは、geocodeによって住所情報と地理的情報がリンクされていることを意味する。

MAF/TIGERに含まれる各レコードは、郵便局(U.S. Postal Service)は郵便物を送付するための住所ファイルとしてDSF(=Delivery Sequence file)と呼ばれるファイルを備えている。そして、郵便物を受け取る人は誰でもDSFに含まれる。郵便局からDSFが無料でアメリカセンサス局に送付される。それによって、LACS(Locatable Address Conversion System)ファイル等がDSFに連結されている。

MAF/TIGERの作成に関する3つの情報源として、上述したDSF、実査による確認(American Community Survey等)、およびLUCA(Local Update of Census Addresses)プログラムを指摘することができる。LUCAプログラムは、州、郡や都市の地方政府当局とセンサス局がパートナーシップをもった場合に可能になるプログラムである。アメリカセンサス局は、地方政府当局にMAFの住所リストを送付し、地方政府当局は住所の追加や削除を行うだけでなく、geocodeを付与する。地方政府当局は地図や境界線を更新し、更新された情報をセンサス局に送付する。それによって、アメリカセンサス局は、MAF/TIGERのデータベースを更新する。なお、近年では、地方・州政府が住所情報を共有することが可能になる地理的支援システム(Geographic Support System=GSS)が開発されている。

つぎに、MAF/TIGERにおける個人情報、以下のように管理される。MAFは、世帯や企業の住所情報を含んでいるが、個別主体の名前は含まれていない。ゆえに、個人が引っ越しても追跡することができないようになっている。また、住所とgeocodeは合、衆国法典第13編(Title 13 of U.S. Code)で保護されている。

1990年までは、アメリカセンサス局は毎回ゼロから住所リストの作成を行っていた。民間部門から住所リストを購入し、リストを確認した上で、欠損している住所を追加する。1990年人口センサスまでは、センサスが終了するたびに、住所リストを廃棄していた。しかし1990年人口センサス以降、住所リストは保管されており、そのリストに新規の住所が追加されている。この住所リストは、現在システム上で管理されている。6か月おきに、郵便局から住所ファイル(DSF)を取得し、1年を通して新規の住所があればファイルに追加されている。

アメリカセンサス局は、税務データといった行政記録情報を利用しようとする場合、行政記録にある住所とMAFの住所データを照合した上で、一意な識別子であるMAF(Master Address File)IDを作成し、行政記録情報に付与している。これによって行政記録情報と人口センサスのマイクロデータに関しては、MAFIDによる連結が可能になる。

さらに、アメリカセンサス局の行政記録研究・分析・利活用センター(Center for Administrative Records Research and Analysis and Applications=CARRA)では、センサス局の担当部局から提供されたMAFを行政記録情報とマッチングした上で、社会保障番号や税務データに関する識別番号が暗号化されたIDである個人識別コード(personal

identification code=PIC)を新たに生成している。MAF ID が住所情報の ID であるのに対して、PIC は、個人レベルでの ID と言える。この PIC を用いることによって、行政記録情報や公的統計データの連結を行うことが可能になる。なお、2010 年センサスの場合、データの 90%に対して個人の識別 ID の付与が可能であることが確認されており、PIC と MAF ID を用いた行政記録情報との連結が可能状況になっている。

PIC は、CARRA とセンサス局の経済研究センター(the Center for Economic Studies)によって管理されている。したがって、例えば、内国歳入庁(Internal Revenue Service=IRS)が管理する税務データがアメリカセンサス局によって入手されると、税務データに PIC と MAF ID が付与される<sup>2</sup>。

アメリカでは、北欧諸国のような人口社会的な情報に関するレジスターは存在しない。そのため、アメリカセンサス局では、行政記録情報の利活用の方向として、北欧諸国のように行政記録に基づくセンサスの作成は、2020 年センサスにおいて指向されていない。2010 年センサスの場合、不在となっていた未回答世帯については、調査員が再訪問を行うという形で再調査を実施していたために、未回答者への再調査(事後調査、follow up)に伴うコストが発生していた。こうしたコストの低減を図ることを目的として、アメリカセンサス局では、2020 年人口センサスにおいて税務データやメディケア(Medicare)等の行政記録情報を活用することが追究されている。図 4 は、アメリカセンサス局における未回答者に再調査に関するモデルを示したものである。未回答者世帯の再調査(Non-response Follow up)のための住所を確認するにあたって、自宅に不在の世帯(vacant)の住所や存在しない世帯(non-existent)の住所を定めるために行政記録が活用されている。行政記録に記載の住所に調査票を郵送し、配達可能だった場合には、未回答者に対して調査対象者から調査票の回答がなされるか、あるいは調査員の訪問による聞き取り調査が行われる。その一方で、行政記録に記載されている住所では調査票が配達不能だった(Undeliverable-As-Addressed)場合には、未回答者は、「行政記録情報において自宅に不在の世帯」かあるいは「行政記録情報において存在しない世帯」に類別され記録される。また、行政記録情報の利用の対象外となった未回答の世帯に対しては、再度の面接調査が実施されるが、未解決の場合に、在室である(occupied)世帯の住所を定めるために行政記録が利用され、行政記録に記載されている住所に調査票が郵送される。それによって、調査対象者からの回答が得られることが期待できるが、あるいは行政記録による補完が行われる場合もある。

このように、アメリカセンサス局では、行政記録情報に用いて未回答者に対する属性情報を把握することが指向されている。このことは、アメリカセンサス局において行政記録情報が統計調査の効率的な実施のための補助情報として機能していることを示唆している。

### 3-5 カナダ統計局における行政記録情報の利活用について

本節では、カナダ統計局を例に、カナダにおける行政記録情報の利活用の状況について見ていきたい。人口センサスデータの作成における行政記録情報の利活用の観点から見た場合、カナダ統計局は、税務データ等の行政記録を用いて人口センサスの統計表の作成可能性を追究している (Lebel and Denis (2016))。カナダ統計局では、住所レジスターを保管しているが、住所レジスターの更新においては、15 種類(2017 年 8 月時点)の行政記録情報が用いられている。具体的には、税務データ、電話サービスの料金明細ファイル(telephone

billing file)、カナダ郵政公社(Canada Post Corporation)の住所ファイル等の最新データを用いて、3か月に1度、住所レジスターが更新されている。そして、人口センサスの実施において、部分的に住所レジスターが利用されている。

カナダ統計局においては、様々な行政記録情報に含まれる社会保険番号(social insurance number=SIN)、運転免許証番号、健康保険証番号等の一意になっている直接的な識別子が、リンケージを可能にするために保管されていることが知られている。さらに、カナダ歳入庁(Canadian Revenue Agency)からの行政記録情報、移民、出生、死亡に関するレジスターの利用の可能性も想定されている。

カナダ統計局における行政記録情報の場合、行政記録情報におけるリンケージに関しては、確率的ではなく、探索的な(heuristic)決定論的なマッチング(deterministic matching)が用いられている<sup>3</sup>。すなわち、名前、住所、出生年、性別、社会保険番号、移住に関する識別子といった直接的な識別子によるマッチングが行われている。

カナダ統計局には、人口レジスター(statistical population register)は創設されていない(2017年8月時点)。しかしながら、カナダ統計局の担当者によれば、個人の統計数値(individual statistical number)が縦断的に連結された構造をイメージしている。これは、カナダ統計局において、人口レジスターを構築するためのレジスターベースの統計システムを指向していると推察される。

カナダ統計局では、学術目的のための行政記録情報の二次利用の例として、Social Data Linkage Environment(SDLE)と呼ばれる二次利用のシステムが展開されている。SDLEとは、レコードリンケージを行うためのセキュアな環境であり、社会人口レジスター(social register)と類似したシステムを備えている。図5は、SDLEの概略図を示したものである。SDLEにおいては、レコードのIDとサーベイデータが含まれず、識別子が保管されたインデックス変数用のファイル(ソースインデックスファイル、Source Index File)、およびレコードのIDと識別子が削除されたサーベイデータが含まれる分析用変数のためのファイル(ソースデータファイル、Source Data File)が別々に保管されている。インデックス変数用のファイルには、性別、出生年月日、性別、郵便番号、社会保険番号といった直接的な識別子が含まれる。さらに、SDLEにおいては、個人の識別子に関する国家レベルの縦断的なファイルである「導出されたレコードの保管庫(Derived Record Depository)」が備わっている。導出されたレコードの保管庫(Derived Record Depository)からキー情報の登録簿(Key Registry)にレコードIDが移管される。このレコードIDは、キー情報の登録簿(Key Registry)には研究に対して必要なレコードのみを発見するためのキーとなる情報として用いられる。行政記録情報のリンケージによって、個人の識別子が含まれない研究に必要な分析用ファイルが生成される。

SDLEにおけるガバナンスのフレームワークとしては、SDLEの担当部局だけでなく、カナダ統計局長(Chief Statistician)とSDLE諮問委員会が、SDLEにおける統計組織的な役割を果たしている。カナダ統計局におけるSDLEプロジェクトでは、申請者は、行政記録情報のリンケージされたデータを用いた研究計画書を提出する必要がある。カナダ統計局長がレコードリンケージに関する研究計画書を検討した上で申請を承認すると、その調査研究は、統計局長から容認された研究として位置づけられる。レコードリンケージの実施のための環境において、リンケージの対象となるデータ源からデータを抽出した上で、ID

だけでなく、ソースインデックスファイルに含まれる直接的な識別子も用いながら、行政記録情報のレコードリンケージが実施される。つぎに、リンクされた分析ファイルを生成するための環境において、レコード ID を用いて、サーベイデータとのリンケージを行われた上で、個人の識別子を削除することによって、容認された調査研究に対してのみ作成される研究に必要な分析用ファイルが生成される。

リンケージが施された分析用の非識別データは、非識別データの利用のためのセキュアな環境を備えるリサーチデータセンターのみでアクセスすることが可能である。こうしたリンクされたデータに関しては、一般公開用型マイクロデータ(Public Use Microdata File)は作成されておらず、あくまで学術研究目的のために利用することが可能になっている。

#### 4. おわりに

本稿では、海外における公的統計マイクロデータの作成・提供および行政記録情報の利活用の動向について論じた。本稿で明らかになったように、公的統計のマイクロデータについては、各国の法制度や利用者のニーズによってマイクロデータの提供形態が異なっている。その場合、統計作成部局と利用者との間に信頼関係が存在することを想定するか(信頼モデル)、あるいはそうした信頼関係を想定しないか(非信頼モデル)によって、個票データのアクセスの可能性は異なる。1つの考え方としては、マイクロデータの提供形態として研究室で個票データの閲覧が可能なりモートアクセスが存在するかどうか、信頼モデルか非信頼モデルかを議論する目安となりうる。公的統計の個票データや行政記録情報のリモートアクセスの仕組みが存在するイギリスやデンマークといった国々については、前者の信頼モデルが当てはまると言える。それに対して、ドイツの場合には、リモートアクセスが法制度的に認められていないことから、後者の非信頼モデルに該当すると思われる。わが国の場合、調査票情報のリモートアクセスの仕組みは現段階では存在しないが、それは、リモートアクセスを可能にするための法制度が存在しないことを意味しない。したがって、わが国の統計法制度のもとで、リモートアクセスの仕組みを可能にするためには、調査票情報の提供に係る統計法 33 条の解釈等、統計法制度の中でリモートアクセスをどのように位置付けるか、また、リモートアクセスに関して統計作成部局と利用者との間の信頼関係をどのように考えていくかを検討する必要があるだろう。

つぎに、匿名化マイクロデータについては、匿名化マイクロデータのニーズによってもその作成方法や提供の仕方は異なってくる。それは具体的には、匿名化マイクロデータを学術研究目的で作成・提供するか、あるいは教育目的を指向するかに依拠する。ニーズに応じて、わが国において今後匿名データ作成の対象となる統計調査や匿名データ作成のために適用される匿名化技法も変わってくるだろう。また、教育目的を指向するのであれば、わが国の統計法制度における可能性を検討した上で、一般公開型マイクロデータの作成可能性を模索することも考えられる。

公的統計データにおけるリモートエグゼキューションについては、ドイツのようにリモートエグゼキューションの仕組みがあっても、その運用については積極的でない国が存在する。その中で、ノルウェーのリモートエグゼキューションシステムについては注目すべき動きであると言える。海外におけるリモートエグゼキューション、さらにはオンデマンド集



計システムの動向については、その可能性も含め今後も注視していく必要があると考える。

本稿では、アメリカセンサス局とカナダ統計局における行政記録情報の利活用の現状も明らかにした。アメリカやカナダは、レジスターベースではなく、統計調査を実施することによってセンサス統計を作成している国々である。とくに、カナダは人口センサスの実施において、インターネットによる回答の促進を図ってきた。アメリカやカナダにおいて、行政記録情報に対する取り組み状況や利活用の方向性は異なっているが、統計作成部局において行政記録情報の利用が積極的に展開されようとしていることは、非常に興味深いと考える。

わが国では、研究者による学術目的のための利用を含む、行政記録情報の利活用に対する関心は高いものの、その利活用が現在十分に展開されているとは言えない。これについては、現行の統計法が、あくまで調査統計を対象にした統計法であることもその一因であるように思われる。本稿で明らかにしたように、レジスターベースの統計作成を行う北欧諸国だけでなく、調査票ベースの統計作成システムを有するイギリスやカナダのような国においても、行政記録情報の二次利用を展開するための制度的な仕組みが存在することを勘案すると、わが国でも行政記録情報の二次利用が可能な形で統計法制度の整備を図っていくことが求められよう。そのために、わが国でも行政記録情報の二次利用にあたっては、現行統計法における行政記録情報の位置付けを明らかにするだけでなく、行政機関等個人情報保護法と統計法との関連性を模索する必要があるだろう。

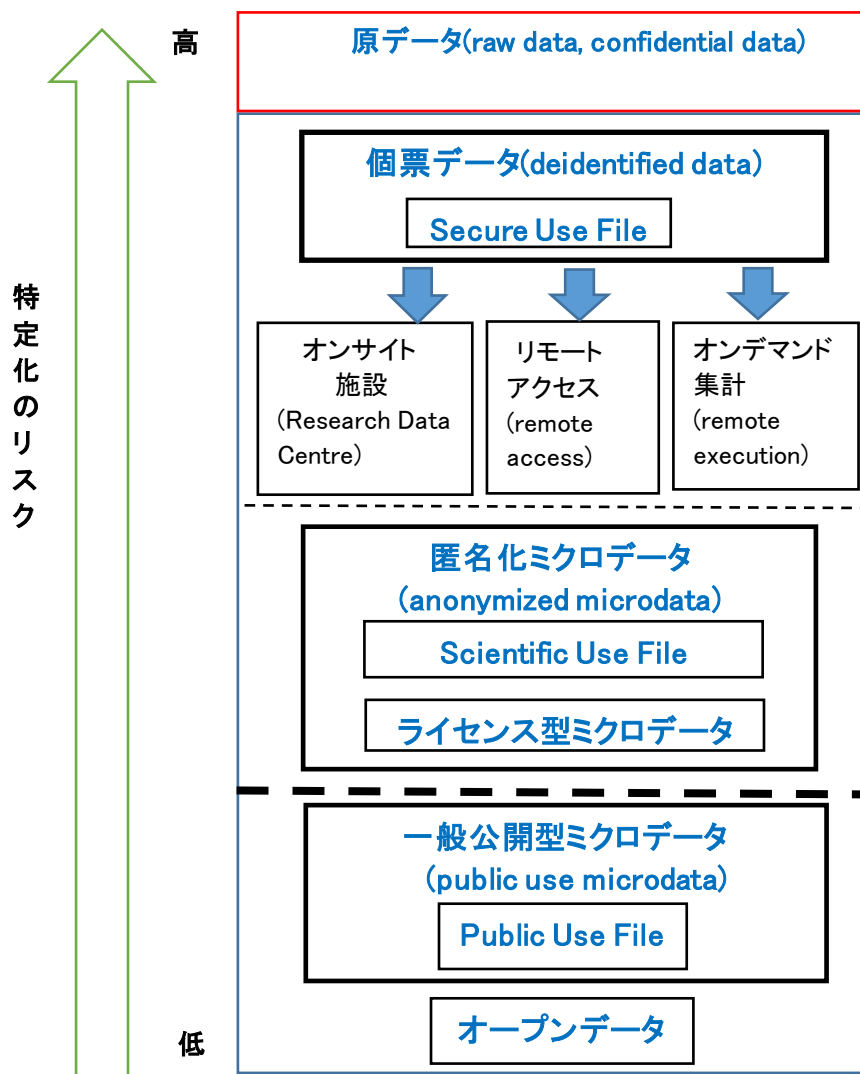
わが国において行政記録情報のさらなる利活用を図る上では、行政記録情報のリンケージを可能にするためのデータの整備も必要である。そのためには、行政記録情報を対象にした共通となる ID の設定可能性とリンケージの方法論が追究することが求められる。なお、アメリカセンサス局のような統計調査における調査客体を捕捉するための住所レジスターとしての行政記録情報の利用可能性、さらには、カナダ統計局で議論されている人口レジスターに向けた行政記録情報の整備も、公的統計の作成における将来的な方向性として検討に値するように思われる。

## 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, 2～9 頁
- 伊藤伸介(2012)「政府統計マイクロデータの提供における匿名化措置—イギリス統計法における法制度的措置と攪乱的手法の適用可能性を中心に—」明海大学『経済学論集』Vol.24, No.3, 1～14 頁
- 伊藤伸介(2014)「イギリスにおける政府統計データの二次的利用の現状」、『ESTRELA』No.241,10～20 頁
- 伊藤伸介(2016a)「わが国における政府統計のデータシェアリングの現状と課題」『情報管理』, Vol.58, No.11, 836～843 頁
- 伊藤伸介(2016b)「政府統計における個票データの提供と秘密保護について—イギリスを例に—」, 『経済学論集(中央大学)』第 56 巻第 5・6 合併号, 1～19 頁
- 伊藤伸介(2016c)「諸外国における政府統計マイクロデータの提供の現状とわが国の課題」,

- 『中央大学経済研究所年報』第48号, 233~249頁
- 伊藤伸介(2017)「公的統計における行政記録データの利活用について—デンマーク, オランダとイギリスの現状—」『経済学論纂(中央大学)』第58巻第1号, 1~17頁
- 伊藤伸介(2018a)「公的統計マイクロデータの利活用における匿名化措置のあり方について」『日本統計学会誌』第47巻第2号, 77~101頁
- 伊藤伸介(2018b)「公的統計マイクロデータの利活用の動向とわが国における課題」『統計』2018年6月号, 13~18頁
- 伊藤伸介・谷道正太郎・小島健一(2018)「オーストラリアにおける公的統計の二次的利用について—オンデマンド集計システム TableBuilder を中心に—」, 『経済学論纂(中央大学)』第58巻第2号, 187~208頁
- 伊藤伸介・木村映善(2019)「医療情報の二次利用の制度に関する現状—デンマークとオランダを例に—」第39回医療情報学連合大会・第20回日本医療情報学会学術大会詳細抄録
- 伊藤伸介・前田幸男(2019)「分野横断的なデータカタログの整備に向けて—現状と課題—」『ESTRELA』No.328, 8~14頁
- 小林良行(2011)「匿名データの教育目的利用に関する一考察」『統計学』第100号, 100~105頁
- 濱砂敬郎(1999)「ドイツ連邦統計法におけるマイクロデータ規定の匿名化措置」, 法政大学日本統計研究所『研究所報』No.25, 69~99頁
- 森博美(2007)「合衆国における人口センサスの新展開」『研究所報』29~48頁
- Heldal, J., Johansen, S., Risnes, Ø. (2019) “Instant Access to Microdata – microdata.no”, Paper Presented at New Techniques and Technologies for Statistics 2019, Brussels.
- Lebel, A. and Denis, J. (2016) “Assessing the usability of a statistical population register for the Census of Population in Canada”, Paper Presented at the United Nations Economic Commission for Europe, Conference of European Statisticians, Group of Experts on Population and Housing Censuses, Geneva, pp.1-16.
- 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.
- Ritchie, F. (2008) “Secure Access to Confidential Microdata: Four Years of the Virtual Microdata Laboratory”, *Economic & Labour Market Review*, Vol.2, No.5, pp.29-34.
- 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.

図1 公的統計の分野における個票データ、匿名化マイクロデータ、Public Use File の位置付け



出所 伊藤(2018b, 14 頁)を一部修正

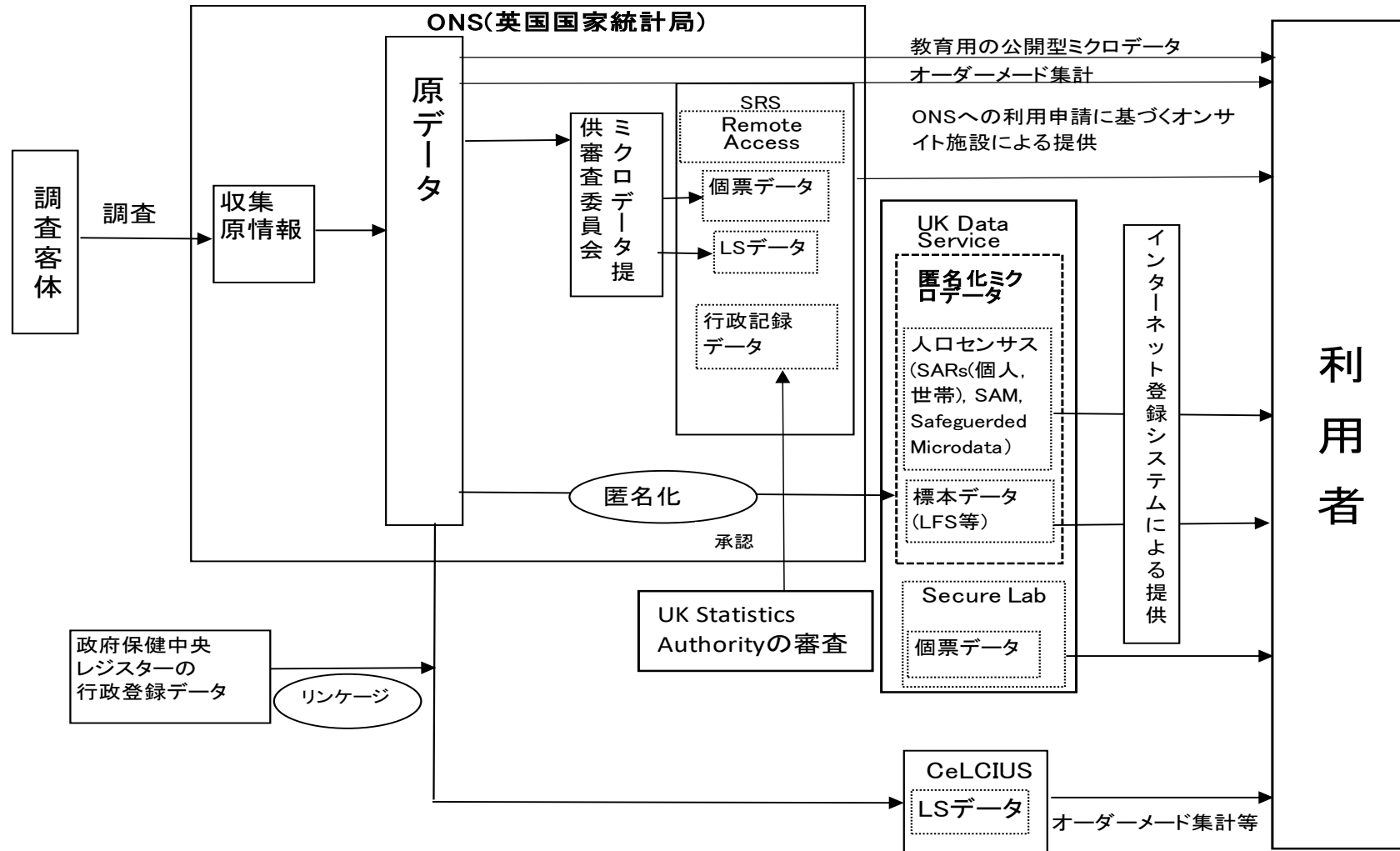
表1 海外の主な統計作成部局におけるマイクロデータの提供形態

データの種類/提供形態	個票データ(非識別データ)の オンサイト施設による提供	個票データ(非識別データ)の 磁気媒体による提供	個票データ(非識別データ)の リモートアクセスによる提供	プログラム送付型のリモートエグ ゼキューションによる提供	オンデマンドシステムによる 提供	匿名化マイクロデータの提供	一般公開型マイクロデータに よる提供
Eurostat	○				●	○	○
イギリス国家統計局	○		○		●	○(UKDSから提供)	○
ドイツ連邦統計局	○			○	○	○	○
フランスINSEE			○			○	○
オランダ統計局	○		○	△	○	○	○
デンマーク統計局			○		○		
フィンランド統計局	△		○			○	
アメリカセンサス局	○				○		○
カナダ統計局	○			○	○		○
オーストラリア統計局	○		○	△	○	○	
ノルウェー統計局	○	○		○			○(NSDから提供)
総務省統計局	○	○				○	

注 ○・・・運営されている。 ●・・・計画中である。△・・・運営中であるが、活動を休止している(2020年2月時点)。

出所 赤谷・荒川・伊藤(2014)、伊藤(2016c)、伊藤(2017)、伊藤(2018a)、伊藤・谷道・小島(2018)、小林(2011)、Zayatz(2007)に基づき作成

図2 イギリスにおけるマイクロデータの提供形態に関する概略図



伊藤(2012), 伊藤(2014), 伊藤(2016b)をもとに作成

図3 ノルウェーにおけるマイクロデータの提供のタイプと有用性・秘匿性の関係

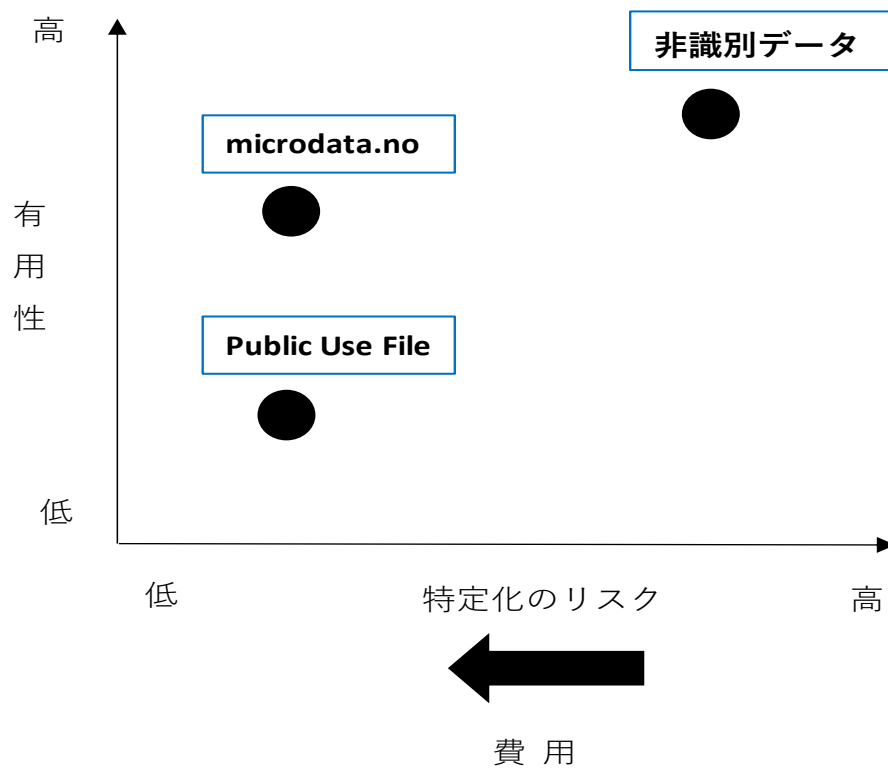


図4 アメリカセンサス局における非回答者の再調査に関するモデル

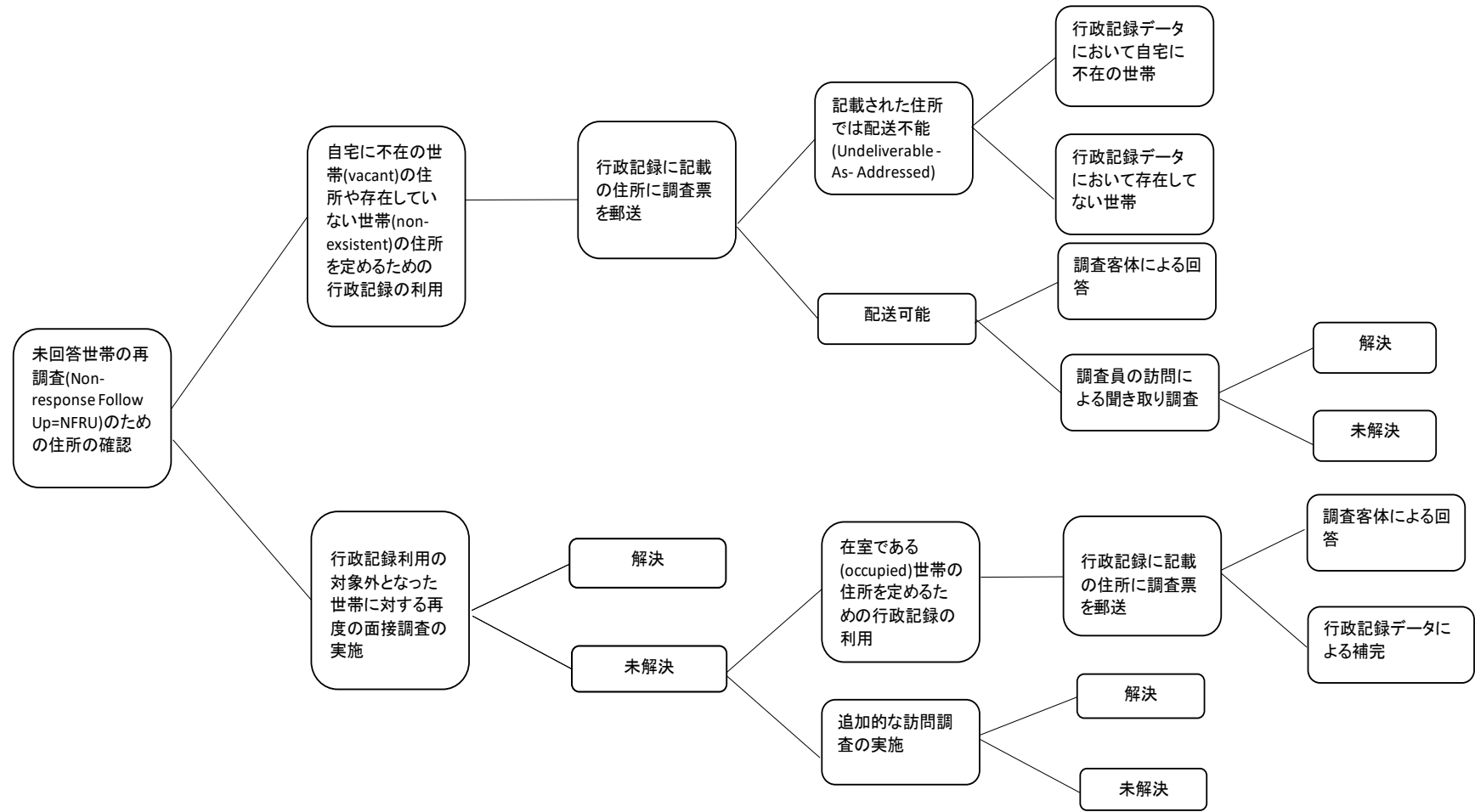


図5 SDLEの概略図

