

## イギリスにおける大規模データの二次利用の新たな展開 —スコットランドの事例を中心に—

伊藤 伸介<sup>†</sup>  
木村 映善<sup>††</sup>  
佐々木 香織<sup>†</sup>

### Recent Developments in Secondary Use of Large Datasets in the United Kingdom: Focusing on the Example of Scotland

ITO Shinsuke  
KIMURA Eizen  
SASAKI Kaori

諸外国では、学術研究目的のために、公的統計マイクロデータだけでなく、医療健康データを含む、行政記録情報の二次利用が広範に展開されている。これらのデータについてはリンケージを可能にするための法制度の整備も進められてきた。本稿では最初に、わが国における公的統計や行政記録情報における二次利用の現状を述べる。つぎに、海外における学術研究目的のための公的統計の二次利用や行政記録情報の利活用が進展している状況を明らかにする。具体的には、Digital Economy Act に基づく法整備やADR UK の支援、さらにはマイクロデータ提供に関する方法的な枠組としてのFive Safes Model のもとで、イギリスで展開されている大規模データの利活用の方向性を確認するだけでなく、スコットランドを事例として、主として行政記録情報を対象にしたデータリンケージの特徴を論じる。最後に、わが国における公的統計や行政記録情報の二次利用における今後の課題を提示する。

キーワード: 公的統計マイクロデータ、行政記録情報、医療健康データ、ADR UK、データリンケージ、スコットランド、リモートアクセス

In many countries, the secondary use of administrative data including medical and health data in combination with official microdata has become standard, and legislation and systems that allow linkage between these data has been established. This paper first investigates the current situation of secondary use of official statistics and administrative data in Japan, and outlines the situation of developing secondary use of official statistics and use of administrative data for scientific purposes. This paper identifies trends in the use of large datasets under the legislation of Digital Economy Act 2017, the support of ADR UK, and the adoption of the Five Safes Model as methodological framework to release microdata. This paper also discusses the characteristics of data linkage for administrative data mainly based on the example of Scotland, and suggests the future tasks to broaden the secondary use of official statistics and administrative data in Japan.

Keywords: Official Microdata, Administrative Data, Medical and Health Data, ADR UK, Data Linkage, Scotland, Remote Access

<sup>†</sup> 中央大学経済学部 Email: ssitoh@tamacc.chuo-u.ac.jp

<sup>††</sup> 愛媛大学医学部 Email: ekimura@m.ehime-u.ac.jp

<sup>†</sup> 北海道公立大学法人札幌医科大学 kaorisasaki@sapmed.ac.jp

## 1. はじめに—公的統計や行政記録情報における二次利用の新たな動き—

近年わが国においては、学術研究目的のための行政記録情報の利活用に対する社会的関心が高まっている。学術研究目的のために、厚生労働省が整備している匿名医療保険等関連情報データベース(NDB: National DataBase)からのレセプト情報等の提供等、医療健康データの二次利用が進展している。また、国税庁は、わが国の税・財政施策の改善・充実等に資する統計的研究を実施するために、2021年6月より国税庁保有行政記録情報(税務データ)の個票データを用いた税務大学校との共同研究を開始した。本共同研究では、特定の研究プロジェクトが当該共同研究に関する公募に採択されると、税務大学校の職員に任用された上で、税務大学校にあるスタンドアローンのPCによる個票データを用いた研究が可能になっている。

それに対して、公的統計の新たな動きとして注目されるのは、規制改革推進会議の第7回医療・介護・感染症対策ワーキング・グループ(以下「WG」と略称)(2023年3月6日開催)において、「公的統計の調査票情報の二次利用について」というテーマで、公的統計の調査票情報の提供に関する今後のあり方について議論がなされたことである。WGにおける議論に基づき、2023年6月1日に「規制改革推進に関する答申～転換期におけるイノベーション・成長の起点～」(以下「答申」と呼称)が発表された。「答申」においては、「統計所管府省庁が利用申出から調査票情報の提供までに要する期間を、令和5年度中に平均1か月以内、令和6年度中に平均1週間以内、かつ、遅くとも4週間に短縮」することが明記された。また、「答申」では、大学・研究機関の研究室等からのリモートアクセスの実証実験の開始とリモートアクセスによる調査票情報の提供の実現にも言及している。このようにわが国では、「答申」を踏まえた形で、調査票情報(個票データ)の提供に関する議論が進められていることから、公的統計の二次的利用のさらなる展開が図られることが期待される。

一方、海外においては、学術研究目的のために、公的統計の個票データだけでなく、行政記録情報さらには医療健康データの利活用が広範に進められてきた(伊藤(2020a, 2020b), 木村他(2020), 佐々木他(2020))。具体的には、社会人口的属性や経済的属性に関する公的統計のマイクロデータを医療・健康・福祉等の他分野とのマイクロデータとリンケージを行うことによって、実証分析の対象範囲が広がってきた。また、各国の統計法制度に基づきながら、攪乱的手法を積極的に適用する形で、公的統計データに対する様々な秘匿措置が追究され、その実用化が図られてきた(伊藤・寺田(2023))。

わが国においては、北欧諸国のように、個人に付与された個体識別番号である共通のIDに基づく公的統計や行政記録情報のリンケージを可能にする仕組み(木村他(2020), 伊藤(2020b))が存在しない。それに対して、わが国と同様に共通IDがないイギリスにおいて、各種の行政記録情報に関するリンケージとそのための法整備が展開されてきたことは、興味深いと言える。このように、海外における大規模データの利活用とそのための秘匿措置の現状を把握することは、わが国における公的統計マイクロデータや行政記録情報の提供における今後の方針を具体的に議論する上でも有益であると考えられる。

本稿では、海外の事例としてイギリスを対象とした上で、公的統計や行政記録情報の二次利用の現状を明らかにする。つぎに、イギリスにおける公的統計や行政記録情報を対象にした大規模データの利活用の方向の1つとして、スコットランドを例に、公的統計マイクロデータや行政記録情報、さらに医療健康データも対象にしたデータリンケージの方法的な特徴を考察する。また、これらのリンケージを可能にしている法制度やFive Safes Modelと呼ばれるマイクロデータ提供に関する方法的な枠組についても論じる。最後に、わが国における公的統計や行政記録情報の二次利用における今後の課題について述べる。

## 2. 欧米諸国における公的統計と行政記録情報の二次利用の現状—イギリスを中心に—

欧米諸国では、公的統計マイクロデータが様々なチャンネルで提供されてきただけでなく、学術研究目的のために行政記録情報や医療健康データの提供が進められた。また、そのための技術的かつ法制度的な秘匿措置も展開されてきた。本節では、イギリスを中心に、公的統計マイクロデータと行政記録情報に関する二次利用に関する最近の状況を明らかにする。

### 2.1 欧米諸国におけるマイクロデータ提供に関する特徴

欧米諸国では、各国の法制度に基づきながら、個体が特定されないように秘匿措置を施した上で、マイクロデータ<sup>1</sup>の提供が行われてきた。提供においては、マイクロデータに対する秘匿性の程度(強度)と利用者の各種のニーズ(有用性)を考慮し、利用者の利用目的も勘案した上で、粒度(詳細さの程度)が異なる各種のマイクロデータが作成され、様々な提供形態を通じて、マイクロデータの利活用の促進が図られてきた。マイクロデータの提供形態は、具体的には、以下のように類型化される(伊藤(2018, 2022))。

第1は、個票データ(非識別データ、**deidentified data**)に対する利用サービスである。個票データについては、学術研究を指向する特定の利用者を対象に、原データ(**raw data, confidential data**)から直接的な識別子を削除するが、追加的な匿名化が施されない研究用の非識別化されたデータとして、アクセスすることが可能である。個票データの提供形態としては、(1)Research Data Centre(RDC)におけるオンサイト施設での利用と(2)リモートエグゼキューション(**remote execution**)における個票データの利用に大別されるが、後者はさらに、大学・研究機関の研究室等からのリモートアクセス、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分析結果の提供、およびオンデマンド集計(リモート集計)による分析結果の提供に類別される。

表1は、2023年9月時点の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を示したものである。表1においては、2020年2月時点の情報に基づき、一部の国々については最新情報への更新を行っている。個票データ(非識別データ)のオンサイト施設による提供を見ると、多くの欧米諸国の統計作成部局が、大学・研究機関さらには統計作成部局の内部に設置されているオンサイト施設による個票データの提供システムを備えていることが確認できる。それに対して、表1で注目すべき点は、欧米諸国における統計作成部局の多くが、大学や研究室等からの個票データのリモートアクセスによる利用サービスを積極的に展開していることである。その中には、デンマークのように、個票データをリモートアクセスでのみ提供している国もある(伊藤(2020b))。最近では、Eurtostat(欧州連合統計局)も、事業所・企業系の統計調査に関して作成された、**Secure Use File**<sup>2</sup>と呼ばれる個票データを対象に、Eurostat内部のオンサイト施設で提供サービスを行うだけでなく(伊藤(2018))、2022年より個票データのリモートアクセスの運用を開始した。さらに、ドイツやカナダのようなリモートアクセスの仕組みを有していない国々においても、リモートアクセスの本格的な実用化に向

<sup>1</sup> 本稿におけるマイクロデータは、公的統計における記入済みの調査個票だけでなく、行政記録情報や医療健康データを対象に個別主体に関して捉えられた個人情報、複数の属性(変数)とそれに対応する属性値(変数値)で編成したレコード群を示している。

<sup>2</sup> Eurostatでは、事業所・企業系の統計調査の場合、個票データに匿名化処理が施された学術研究用ファイル(**Scientific Use File**)は作成されておらず、**Secure Use File**と呼ばれる匿名化技法が適用されていない個票データのみが利用可能になっている。なお、Eurostatにおいて**Secure Use File**が提供されている統計調査は、**Community Innovation Survey**と**Structure of Earnings Survey**の2調査である。

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 オンライン施設による提供	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
各国の統計作成部局							
Eurostat	○		○			○	○
イギリス国家統計局	○		○		○	○(UKDSから提供)	○
ドイツ連邦統計局	○		●(SUFとして提供)	○	○	○	○
フランスINSEE			○			○	○
オランダ統計局	○		○	△	○	○	○
デンマーク統計局			○		○		
スウェーデン統計局	△		○			○	
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：UKDSとはイギリスエセックス大学のUK Data Service、NSDとはNorwegian Centre for Research Dataの略である。

○・・・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月時点の情報をもとに、一部の国々については2023年9月時点の情報に更新)。

出所 伊藤(2020a), 表1を一部修正

けた動きが展開されている点が興味深い<sup>3</sup>。なお、ドイツやノルウェーのように、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を行っている国や、オーストラリアのように、利用者のニーズに応じた形で、個票データからオンデマンドで作成した集計表を利用者に提供するTable Builderというオンデマンド集計システム(伊藤他(2018))を有する国も存在する。

第2は、匿名化されたマイクロデータの提供である。これについても、学術研究目的のために作成される匿名化マイクロデータ(個票データに匿名化処理が施されたデータ、anonymized microdata)の作成・提供、および一般には利用目的の制限がなく誰でも入手可能な一般公開型マイクロデータ(public use microdata)の作成・公開に大別することができる。前者については、

<sup>3</sup> Eurostatの場合、2022年より、Eurostatの外部の限定された大学・研究機関において(デンマークとイタリアの2カ所)、リモートでEurostatが保有する公的統計の個票データのアクセスが可能になっている(Bujnowska et al.(2023))。また、ドイツやカナダにおいても、リモートアクセスの本格的な準備を進めている。ドイツにおいては、ドイツ連邦統計局が、事実上の匿名性(factual anonymity)の概念にしたがって作成された学術研究用ファイル(Scientific Use File=SUF)をより粒度を細かくしたSUFを研究室からリモートでアクセスするシステムを整備する動きがある(Brenzel et al.(2023))。カナダにおいても、カナダ統計局がリモートアクセスの実用化に向けて準備を進めている。カナダ統計局におけるリモートアクセスシステムであるVirtual Data Labプロジェクトの検討状況に関しては、Cranwick(2019)および下記のURLを参照。

<https://www.statcan.gc.ca/en/about/er/vdplp>

表1では、イギリス、ドイツ、Eurostatのように学術研究用ファイル(Scientific Use File=SUF)の形で提供されていることが確認できる<sup>4</sup>。SUFの利用においては利用申請が必要であるだけでなく、利用のためのライセンス(誓約書)が求められるケースが少なくない。また、後者においては、欧米における多くの国々で、一般公開型ファイル(Public Use File=PUF)としての提供が行われている。SUFとは異なり、PUFの利用目的に関する制約はないが、教育目的やプログラム作成のための利用が推奨されていることは、留意すべき点と思われる。

## 2.2 イギリスにおけるマイクロデータの提供状況

本節では、イギリスにおけるマイクロデータ提供の現状について概括する。イギリスでも同様に、①個票データの利用サービス、②匿名化マイクロデータの提供、および③一般公開型マイクロデータの公開という形で、マイクロデータの提供が展開されてきた。イギリスのマイクロデータに関する多様な提供形態について議論することは、わが国における公的統計の二次利用の展開方向を模索する上でも参考になると考える。

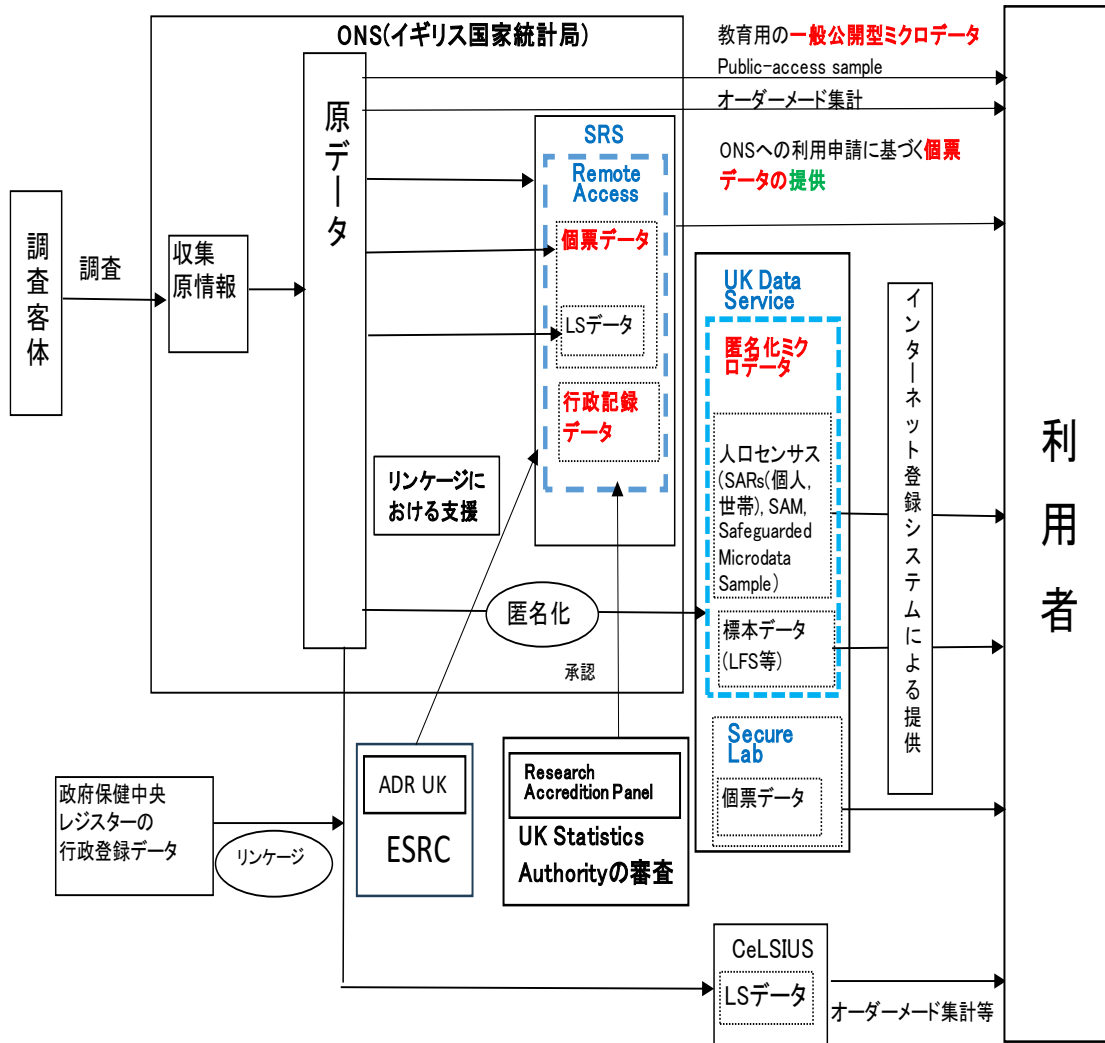
図1は、イギリスにおけるマイクロデータの提供形態に関する概略図を示したものである。統計作成部局であるイギリス国家統計局(Office for National Statistics=ONS)では、公的統計の個票データと行政記録情報の提供を担当する部署である Secure Research Service(SRS)が、オンサイト施設を運営しており、研究者は SRS を通じて個票データの利用が可能になっている。その中には、CeLSIUS(=Centre for Longitudinal Study Information and User Support)が提供の窓口になっているセンサスレコードの縦断的なリンケージデータである ONS 縦断データ(ONS Longitudinal Study、以下「LS データ」と呼称)のオンサイト利用も含まれる。また、ONS の SRS が管理するオンサイト施設だけでなく、エセックス大学に付置されたデータサービス提供センターである UK Data Service(UKDS)が管理・運営する the Secure Lab を通じて、研究者は、大学・研究機関の研究室から個票データをリモートでアクセスすることも可能である。

これに関する注目すべき点は、イギリスでは、公的統計の個票データだけでなく、行政記録情報と公的統計マイクロデータがリンケージされた行政記録情報、および行政記録情報同士でリンケージされたデータの利活用が広範に展開されていることである。これについては、後述する2018年に施行された Digital Economy Act 2017(以下、DEA と略称)と ADR UK の支援による行政記録情報のリンケージサービスを指摘することができる。ADR UK とは、イギリスの研究者に対する学術研究のための資金提供機関である ESRC(=Economic Social Research Council)によって運営される行政記録情報の利活用の推進に関するプロジェクトである。それ以前は、ESRC が、行政記録情報のリンケージを推進するために、Administrative Data Research Network と呼ばれるプロジェクトを展開していたが、これについては第3節で詳述する。

一方、イギリスでは、UKDS において匿名化マイクロデータが提供されている。イギリスにおいては、匿名化マイクロデータの安全な利用に関する誓約書を提出し、ライセンス(End User Licence)を取得すれば、UKDS のサイトで ID とパスワードを入力することによって、匿名化マイクロデータをダウンロードして利用することができる。以下では、イギリスで提供されているライセンスを要する匿名化マイクロデータを「ライセンス型マイクロデータ」と呼称する。

<sup>4</sup> 近年、欧米諸国では、オンサイト施設やリモートアクセスのような安全な環境の下における個票データの提供が広範に展開されているが、Eurostatのように、利用申請の多くがSUFとなるようなケースもある。

図1 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



注 SARs: Samples of Anonymised Records, SAM: Small Area Microdata, CeLSIUS: Centre for Longitudinal Study Information and User Support

出所 伊藤(2020a) 図2 を一部修正

ライセンス型マイクロデータの例としては、人口センサスの匿名化マイクロデータ(1991年と2001年の人口センサスのSARs(=Sample of Anonymised Records)、2001年センサスのSAM(=Small Area Microdata)、2011年センサスのSafeguarded microdata sample等)や労働力調査のような標本調査データ(サーベイマイクロデータ)が指摘できる。2021年の人口センサスでは、Safeguarded household sampleという世帯単位で抽出され、階層構造を持つライセンス型マイクロデータの作成・提供が計画されている。また、一般公開型マイクロデータとしては、2011年センサスに関する教育用のマイクロデータが、ONSのウェブサイトで開催されている。なお、2021年センサスに関しても同様に、ONSがpublic-access sampleを作成・公開することが計画されている(伊藤・寺田(2023))。

このような、イギリスにおける公的統計マイクロデータや行政記録情報に関する現在の提供形態については、ONSによって考案され、現在ではヨーロッパ諸国だけでなく、カナダやオーストラリアといった国々でも広範に展開されている「5つの安全モデル(Five Safes Model)

が、適用されたものと考えることができる。5つの安全モデルは、当時イギリス国家統計局の職員であった Felix Ritchie 博士(現在、西イングランド大学ブリストル校(University of the West of England Bristol=UWE Bristol)教授)によって 2003 年に考案されたモデルである((Ritchie(2008), Desai et al.(2016), 伊藤(2016b))。それは、マイクロデータ提供を対象にした統計実務の方法論的な基礎となるデータ提供の法的制度的な枠組みであって、主としてヨーロッパ諸国における公的統計マイクロデータや行政記録情報だけでなく、近年では医療健康データの安全な利活用を図るための有力なモデルとしても展開されている(伊藤(2016b), 伊藤(2020a))。

5つの安全モデルは、以下の5つの基準から成るマイクロデータの提供に関する実践的なモデルとして位置付けられる。すなわちそれは、①妥当な統計目的のために個票データの利用を行うプロジェクトと判断される安全なプロジェクト(safe projects)、②個票データの利用にあたって信頼できる研究者であると認可された安全な利用者(safe people)、③個人情報が見えないデータであることを示す安全なデータ(safe data)、④個票データの利用にあたっての技術的な安全管理措置が施された安全な施設(safe settings)、および⑤個人情報の露見リスクを有する分析結果が含まれていない安全な分析結果(safe outputs)である(伊藤(2016b), 伊藤(2018))。

5つの安全モデルの考え方によれば、これらのひとつひとつの基準を組み合わせ、「安全な統計(safe statistics)」となるレベルに達するまで安全性の強度を高めることによって、安全な統計の作成・提供が可能になる。例えば、統計作成部局が個票データを提供する上では、安全なプロジェクト、安全な利用者、安全な施設と安全な分析結果を重要視する。これらの4つの安全性を確保した上で、安全なデータについては、個票データは、個体に関する直接的な識別子の削除のみが適用された非識別データが、オンサイト施設やリモートアクセスのような安全な施設の中で可能になる。対照的に、一般公開型マイクロデータの場合には、ウェブで誰でも入手可能なマイクロデータを公開する上で、安全なプロジェクト、安全な利用者、安全な施設と安全な分析結果は要求されない。この場合、安全なデータが重要視され、そのために秘匿性の強度を高める形で、一般化の技法(リコーディング、トップ(ボトム)・コーディング)やデータ削除(レコード(行)の削除、変数(列)の削除、セルの削除)といった非攪乱的手法だけでなく、ノイズ付与、スワッピング等の攪乱的手法が適用される<sup>5</sup>。

### 2.3 Secure Research Service における公的統計の個票データの提供サービス

ONS の SRS(=Secure Research Service)は、「承認された研究者(approved researcher)」あるいは「認可された研究者(accredited researcher)」の資格を取得している利用者に対してのみ、セキュリティの管理がなされた環境のもとで、個票データのアクセスを容認している(伊藤(2016a, 2020a))。「承認された研究者」は、イギリス統計法(The Statistics and Registration Service Act 2007)の「第 39 条 個人情報の秘密保護」(5)(i)の中で定義された概念であって、法第 39 条では、「(法人を含む)特定の個人と関連付けられ、およびその個人を識別する情報」である「個人情報」の秘密保護の対象外に置かれている。すなわち、「承認された研究者」は、「統計研究の目的のために、統計委員会(Statistics Board)によって保有される個人情報を取得することについて委員会が容認した者を意味している」(法 39 条(5))(伊藤(2016a))。それに対して、「認可された研究者」は、2018 年に施行された DEA に基づいて定められた利用者が個票データの利用のために求められる資格である。DEA の「第 64 条 研究目的のための情報の開示 (disclosure)」の第 1 項には、「公的な当局(public authority)によって保有される、公的当局と関

<sup>5</sup> UWE Bristol の Felix Ritchie 教授へのインタビューに基づいている。

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」とあり、この「公的当局と関連している情報」が、行政記録情報に該当することから、DEAは、主として行政機関が保有する行政記録情報および各種の行政記録情報がリンケージされたデータの提供を可能にするための法律として位置付けられている。したがって、DEAに基づいて行政記録情報を利用するにあたっては、利用者は、「認可された研究者」になるための資格を取得することが求められる(伊藤(2020a))。なお、ONSのSRSの担当職員によれば、根拠となる法律は異なるものの、操作上の概念としては「承認された研究者」と「認可された研究者」に実質的な相違はないと言える。さらに、DEAの中の個票データの提供に関わる一部の条文はイギリス統計法と関連付けられているだけでなく、DEAに基づき、行政記録情報間だけでなく公的統計マイクロデータと行政記録情報のリンケージされたデータが申請される現状を勘案すると、利用者が「承認された研究者」よりも「認可された研究者」の資格を取得することが、近年では通常の考え方となっているとのことである。

利用者が、個票データを申請する上での最初の手順は、「承認された研究者」あるいは「認可された研究者」の資格を取得することであるが、この資格を取得するための申請書には、主として以下の事項の記載が求められる。

- ①申請者の所属機関
- ②計量分析を行った経験
- ③統計データ(個票データ)の利用経験
  - (1)利用した統計データ
  - (2)統計的手法
  - (3)利用したソフトウェア
  - (4)分析結果に対する秘匿措置

なお、認可された研究者(承認された研究者)の資格取得のためには、個票データを利用するためのトレーニングも義務付けられている。

認可された研究者としての資格を取得すると、つぎに、実施を計画している研究プロジェクトが認可される必要がある。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、ONSのSRSが利用申請の窓口となって、申請書類をチェックする。その上で、記載内容に問題がなければ、データの処理を行う者(processors)、研究プロジェクトと研究者のそれぞれの安全性について、UK Statistics Authorityの研究認定委員会(Research Accreditation Panel=RAP)(ONSに設定されていたMicrodata Release Panelの後継の審査委員会)が一元的に審査を行う。RAPによる審査を通過すると、申請中の研究プロジェクトが認可され、ONSに設置されているオンサイト施設であるSafe Roomsやセキュリティの管理がなされた環境としてのSafePods、さらにはリモートアクセスの形で個票データの利用が可能になる。この「認可されたプロジェクト」となるための申請書類に関する主な記載事項は、以下のとおりである。

- ①研究プロジェクトの目的や概要
- ②研究プロジェクトの期間
- ③モデル分析の手法<sup>6</sup>

---

<sup>6</sup> 申出書においては、どのような分析を行うのかについては明記する必要があるものの、具体的な統計表やモデルで使用する変数については明示する必要はない。



- ④利用する個票データの名称、年次
- ⑤使用するソフトウェア等

審査における採択の可否は、申請対象となる研究が公共の利益(public good)をもたらすかどうかに基づいて検討され、採択された場合に個票データのアクセスが行われる。

申請手続きの効率性といった観点から見れば、個票データの利用申請において、分析するために必要な変数を事前に提示することが要件となっているか否かは、重要な論点だと言える。イギリスの場合、申請書類における利用する変数の提示は、利用する個票データのタイプ、およびどのような目的で個票データを利用するかによるという点で、「ケースバイケース」だと言える。例えば、標本調査の個票データの申請の場合、利用する変数の選択が必要なく、事前に必要な変数の提示が求められないケースが存在する。それに対して、人口センサスのような大規模なデータについては、個票データの利用において多様な実証研究の方向が想定されることから、利用する変数を事前に提示することが求められている。SRSの担当者によれば、行政記録情報とのリンケージデータについても、研究者は事前に必要な変数を示すことを求められるとのことである。さらに、縦断的なデータ構造を備えるLSデータ(ONS Longitudinal Study)のように、追加的な申請書類の提出が必要であり、LSデータを申請する理由や分析に使用する変数の一覧表の記載の提示が求められる事例もある。

SRSにおいて利用可能な個票データは2023年6月時点で125のデータセットである。の中で、行政記録情報に関しては、教育データ、裁判データ、および税務データが、利用申請が多い行政記録情報となっている。これらの大部分においては、DEAに基づいて学術研究のための利用が可能になっている<sup>7</sup>。さらに、氏名や住所のような直接的な識別子を用いてONS内部でリンケージが行われた公的統計の個票データに関しても、SRSを通じてアクセスすることが可能である。ただし、SRSでは、このような識別子にアクセスすることはできず、非識別データのみアクセスすることが可能になっている。ONS内部におけるデータリンケージの方法については第3節で述べる。なお、リンケージ用に生成されたIDを用いれば、研究者自身がSRS内でマッチングを行うことも可能とのことである。

申請者は、利用申請にあたって、個票データに関するメタデータを確認した上で、必要な変数を選定する。現在ONSでは、Secure Research Service Metadata Catalogueと呼ばれる個票データに関するメタデータのカタログのウェブページが整備されており、利用者は、当該ウェブページを通じて、制限なく利用可能な個票データに関する一括検索が可能になっている<sup>8</sup>。興味深い点は、公的統計や行政記録情報がリンケージされたデータも、メタデータカタログにおける検索対象になっていることである。このことは、研究プロジェクトによって使用されていたリンケージデータの保存・共有が、SRS内部でなされていることを示唆している。

イギリスでは、申請者が個票データの利用申請を行うと、そこから個票データの提供までに要する期間がケースバイケースで異なることが確認された。ONSの担当者によれば、イギリスの場合、利用申請から最短では1か月で提供可能になる。具体的には申請者が、認可された研究者(承認された研究者)を取得するために、最短で1週間、研究プロジェクトがUKSAのRAPで認められるまでに最低3週間かかるとのことである。ただし、申請する個票デー

<sup>7</sup> 教育データに関する行政記録情報は、DEAの適用の対象外となっている。

<sup>8</sup> イギリスの公的統計マイクロデータの検索については、UKDSが、公的統計のライセンス型マイクロデータを対象に、社会調査の個票データ等も含む一括検索が可能なデータカタログシステムを整備していたが、ONSが構築したメタデータの検索システムは、公的統計の個票データの検索に特化している点で、UKDSのデータカタログとは大きく異なる。

タの特性に応じて提供される期間は異なるということであった。なお、個票データの利用が容認された場合、平均して1年～1年6か月、個票データがSRSにおいて利用されている。

SRSでは現在、オンサイト施設での個票データの利用だけでなく、リモートでアクセス可能な個票データの研究者に対する提供サービスも広範に展開されている。イギリスでは、2020年に新型コロナウイルス感染症が急速に蔓延する以前から、個票データの新たな利用サービスとして、「保証された組織の接続可能性(Assured Organisational Connectivity=AOC)」と呼ばれる協定をONSとの間で締結した研究機関に所属する研究者を対象に、SRSに訪問してオンサイト施設で分析を行わなくても、所属する研究機関からリモートで個票データにアクセス可能な仕組みが整備された。ただし、利用者はイギリス国内においてのみ、個票データへのリモートアクセスを行うことが可能である。さらに、自宅からのリモートアクセスについても、データの所有者とONSによる承認の可否に従って、ケースバイケースによる対応がなされている<sup>9</sup>。なお、ONSの担当者からのインタビュー調査に基づく注目すべき情報として、個票データにアクセスしている研究者の90%以上が、研究室からのリモートアクセスで、個票データを利用していることが指摘できる。

行政記録情報の利用の認可が得られると、ONSのSRSにおけるオンサイト施設やリモートアクセス施設、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。さらに、行政記録情報を用いて行われた分析結果の持ち出しに関する審査については、公的統計の個票データにおける分析結果の場合と同様のチェック基準が適用されている。具体的には、5つの安全モデルにおける「安全な統計」を獲得するための分析結果のチェックの方法として、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「原則対応型アプローチ(Principles-Based Approach)」が採用されている(伊藤(2020a))<sup>10</sup>。

オンサイト施設やリモートアクセス施設といった安全な施設からの分析結果の持ち出しに関しては、分析で用いたデータのタイプや規模によって、持ち出しに関する基準は異なる。例えば、世帯・人口系の統計調査に関する集計表を持ち出す場合、セルにおける閾値は10に設定されていることから、集計表に含まれるセルの中の度数が10以上であることが求められる。なお、事業所・企業系の調査の場合には、閾値は30となっている。なお、ONSの担当者へのインタビューによれば、近年実証研究が広範に進められている機械学習モデルを用いた分析に関しては、SRSは機械学習の分析結果を持ち出すための特別なチェックを想定していない。また、SRSにおいて、(社会科学の分野において)機械学習モデルにおいて実証研究を行っている研究プロジェクトはないとのことである(2023年6月時点)。

### 3. イギリスにおける大規模データにおけるリンケージの動向

<sup>9</sup> AOCについては、研究室といった職場からのリモートアクセス(AOC Office access)と自宅からのリモートアクセス(AOC Remote access)がある。これらについては、所属長の署名がついたONSとの協定書とセキュリティ管理に関する文書を提出することによって、リモートアクセスが可能になる。

<sup>10</sup> 分析結果に関する最終成果物のチェックに関するアプローチとしては、「ルール対応型アプローチ(Rule-Based Approach)」と「原則対応型アプローチ」の2つがある。前者については、閾値ルール(threshold rule)といった「一連のルールに基づいて、分析結果が公表可能かについての判断を行うこと」にその特徴がある。それに対して、後者における「特徴としては、原則として分析結果の公表の可否は、チェックを行う担当者に委ねられていること」が指摘できる(伊藤(2016b, pp.241-242))。5つの安全モデルの原則的な考え方に立てば、データの提供者と利用者との信頼関係に基づき、分析結果の持ち出しは、「ケースバイケース」で対応することから、原則対応型アプローチが適用される(伊藤(2016b))。それに対して、ドイツのように、オンサイト施設やリモートアクセスのような安全な環境で個票データを分析した後の持ち出し審査において、「ルール対応型アプローチ」が用いられている事例もある。

前節でも述べたように、イギリスでは、DEA が施行されることによって、行政記録情報同士、あるいは行政記録情報と公的統計マイクロデータのリンケージが法制度的に整備されたことから、リンケージデータの利活用が広範に展開されている。それについては、リンケージされた公的統計の個票データおよび行政記録情報を利用した実証研究が、公益性を有する研究(public good research)であることが、イギリスにおいて学術研究のためにリンケージされた行政記録情報を利用するための要件になっている(伊藤(2020a))<sup>11</sup>。こうした状況を踏まえた上で、本節では、イギリスにおける大規模データのリンケージの特徴について論じる。

イギリスでは、リンケージされた行政記録情報の二次利用の展開は、2011年にESRCの主導によって、学術研究や政策評価のための行政記録情報の提供を指向して設置されたAdministrative Data Taskforce(ADT)が契機となっている(伊藤(2016a))。ADTが2012年12月に刊行した報告書『研究と政策のためのデータアクセスの改善に向けて』(Improving Access for Research and Policy)では、「新たな研究機会のための行政記録情報の共有(Sharing Government Administrative Data: new research opportunities)」に関するプロジェクトが提唱され、複数の行政記録情報同士、および公的統計の個票データと行政記録情報においてリンケージされたデータの利活用を可能にするために、Administrative Data Research Network(ADRN)が創設され、2014年10月から行政記録情報の提供サービスが開始された。

ADRNプロジェクトは、エセックス大学において行政記録情報の申請手続きおよび審査を行うAdministrative Data Service(ADS)、およびイングランド、ウェールズ、スコットランドと北アイルランドというイギリスを構成する各国が独立して運営されるAdministrative Data Research Centre(ADRC)として組織化された。利用者は、ADRNによる審査によって、行政記録情報の利用の認可を得ると、ADRC内部のオンサイト施設でリンケージされた行政記録情報の非識別データを利用することができた。

ADRNプロジェクトにおける非識別データは、以下の手順で作成・提供された(伊藤(2016a))。①利用者は、行政記録情報の利用に関する研究計画書を提出する。②研究計画書をもとに、ADRNは当該研究プロジェクトに必要なデータを提供するように、分析の対象となる行政記録情報を所管する行政機関(データの保管者(data custodian))と交渉を行う。③データの提供を承諾した行政機関は、氏名、生年月日等の個人に関する直接識別子(direct identifier)を切り離れた上で、各レコードに一意的な参照番号を付与する。④この参照番号が付与され、直接的な識別情報が削除されたデータがADRCに送付される。⑤行政機関は、直接的な識別情報と一意的な参照番号の組のみを第三者機関に送付する。⑥第三者機関が、リンケージの対象となる複数のデータのそれぞれにおいて同一の個人に付与された各種の一意的な参照番号と直接識別子との間でマッチングを行い、直接識別子を消去した上で、同一の個人に関連付けられた一意的な参照番号の組であるindex keyのみをADRCに送付する。⑦ADRCは、index keyを用いてデータのリンケージを行った後にindex keyを消去する。⑧利用者は、ADRCのオンサイト施設でこのリンケージデータにアクセスして実証研究を行う。

ADRNプロジェクトは2018年に終了したが、その後継プロジェクトが、第2節で述べたADRUKとなる。イギリス全体(UK)を統括するADRUKの下で、イングランド、ウェールズ、スコットランドと北アイルランドのそれぞれにおいて、行政記録情報のリンケージされたデータの利用サービスが個別に展開されている(伊藤(2020a))。イングランドの場合、ONS

<sup>11</sup> イギリスでは、国民健康サービス(National Health Service)が保有・管理を行う医療健康データについては、国民の個人情報保護に対する政府への根強い不信があった状況下であっても、公益性を有する研究という観点から、学術研究目的で医療健康データの利用が可能になった。詳細については、佐々木他(2020)を参照。

が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一元的に担っているが、ADR UK は、ONS とのパートナー関係を結んでおり、資金提供を担う研究インフラ支援機関(funding agency)として、ONS におけるデータリンケージの仕組みを整備するための資金援助を行うだけでなく、行政記録情報を保有する行政機関との交渉も担っている。さらに、ADR UK は、UKSA の RAP による行政記録情報の提供の可否に関する審査についても、審査メンバーの一員として関わっている。

現在、ONS が行っている非識別データの提供方法は以下のようである。

- ① ONS でリンケージを専門に行っている部署(SRS とは異なる)において、10~20 人程度でリンケージの作業を行う。リンケージの対象となる複数のマイクロデータの中の各レコードから個人を直接識別することが可能な情報(氏名、住所、生年月日等)のみ切り離す。
- ② 直接識別子以外の情報は切り落とされた上で、仮名化を行った上で、リンケージ用 ID が生成される。直接識別子のマッチングを行い、マッチングされたリンケージ用 ID の対応表のみを残す。このとき、直接識別子は消去される。
- ③ ONS のデータリンケージ担当チームは、このリンケージ用 ID の対応表に基づいて複数のマイクロデータ間のリンケージを行う。
- ④ リンケージ後に、リンケージ用 ID が削除されたリンケージデータが SRS に保管される。研究者は、SRS のオンサイト施設においてリンケージデータにアクセスするか、研究室からリモートアクセスすることによって、研究を進める。

ONS がデータリンケージで採用している手続きは、ADRN プロジェクトで適用された方法に類似しているようにも思われる。しかしながら、ADRN プロジェクトでは、データリンケージの対象となる行政記録情報を保有する行政機関に対する調整に関しては、ADRC が設置されていたエセックス大学の担当職員が担っていたが、行政記録情報のリンケージに関する調整が困難な状況も存在していたことが推察される。それに対して、行政機関との調整とデータリンケージの実務に関する現在の状況としては、ADR UK と ONS で役割分担がなされているだけでなく、ONS 内部でデータリンケージの手続きが完結していることが指摘できる。その意味では、行政機関との調整、データのリンケージ業務、研究者へのリンケージデータの提供は、一体的に展開されているため、データリンケージの業務にあたって、より効率的な手続きが進められているとすることができる。

なお、特定の研究プロジェクトに基づいて、一度リンケージされたデータは、他の研究プロジェクトも申請すると、リンケージデータを利用することが可能になっている点は、興味深い。その意味で、前節で述べたように、ONS の SRS によって、リンケージデータも含む形で、検索可能なメタデータが整備された点は、研究者にとって、マイクロデータの利用における利便性を高めているとすることができる。

#### 4. スコットランドにおける行政記録情報のリンケージについて

本節では、イギリスのスコットランドを対象にした上で、行政記録情報や医療健康データの二次利用とデータリンケージの現状について論じる。本節で、スコットランドを対象に行政記録情報の利活用について議論する理由は、イングランドとは異なり、スコットランドでは、人口社会および社会経済に関する行政記録情報間のデータリンケージ、あるいは行政記録情報と公的統計マイクロデータとのリンケージだけでなく、行政記録情報と医療健康データとのリンケージも可能にする仕組みが整備されているからである。

スコットランドは、イングランドとは大きく異なるデータ提供体制を有している。それは、国民健康保険サービス(National Health Service, 以下 NHS と呼称)が保有する医療健康データを対象に、学術研究目的のために研究者に対する提供業務を担っている Public Health Scotland(PHS)が、社会経済を対象にした人口センサス等の公的統計マイクロデータや行政記録情報の提供においても主要な役割を果たしていることである。それに関しては、エディンバラ大学に設置されている ADR Scotland が、PHS に資金提供を行っており、このことが、PHS に医療健康データだけでなく、社会経済分野における行政記録情報の提供サービス業務を担当させることを可能にしている。具体的には、PHS 内部に設置されている eDRIS(=electronic Data Research and Innovation Service)が、個票データの申請手続きを行うための窓口になっているだけでなく、行政記録情報間のリンケージやセンサス等の他の公的統計マイクロデータとの連結に関する業務も担っている。さらには、eDRIS によって、行政記録情報と医療健康データのリンケージされたデータを研究者がアクセスすることが可能になっている<sup>12</sup>。これについては、eDRIS が、スコットランドで整備された National Safe Haven(NSH)と呼ばれる、個票データを安全に利用するための「信頼された研究環境(Trusted Research Environment=TRE)」の中に位置付けられる点が指摘される。

TRE は、先述の5つの安全モデルに基づいて設計されている。すなわちそれは、認可された研究者(「安全な利用者」)が、承認された研究プロジェクト(「安全なプロジェクト」)の下で、(スコットランドでは、eDRIS が業務を担っている)データリンケージと非識別化がなされた個票データ(「安全なデータ」)を用いて、安全管理措置がなされた環境(「安全な施設」)の中で分析を行い、そこから秘匿性に関するチェックが行われた「安全な分析結果」のみを持ち出すことが可能な仕組みを備えている<sup>13</sup>。

スコットランドの住民によって登録された、出生、死亡、結婚、離婚の人口動態および世帯変動に関する行政記録情報やスコットランドのセンサスレコードを保管している National Records of Scotland(NRS)も、スコットランドにおけるマイクロデータのリンケージおよびリンケージされたマイクロデータの利活用の進展に重要な役割を果たしている。NRS は、スコットランド総合登録局(the General Register Office for Scotland)とスコットランド国立公文書館(the National Archives of Scotland)が統合されて2011年に設立された、スコットランド政府から独立した行政機関(Non-Ministerial Department)である。NRS も、ADR Scotland とパートナーシップ関係にあるだけでなく、信頼できる第三者機関(Third Trust Party)として、行政記録情報に関するリンケージの手続きに関わっている。また、NRS 自身も、人口センサスや出生・死亡の人口動態情報についてリンケージされたデータの提供を行っており、認可された研究者は、安全な環境において、このリンケージデータを利用した研究を行うことができる。

なお、スコットランドにおける行政記録情報の利用形態については、研究者はリモートアクセスでの利用が可能になっている。また、イングランドに位置するONSが、DEAに基づいて行政記録情報の大半を提供しているのとは異なり、スコットランドの行政記録情報の利

<sup>12</sup> エディンバラ大学のChris Dibben教授は、人口センサスの縦断データ(longitudinal data)に関する実証研究を行うスコットランド縦断研究センター(Longitudinal Studies Centre Scotland)のセンター長であるだけでなく、スコットランド行政記録情報研究センター(Scottish Centre for Administrative Data Research)およびADR Scotlandのセンター長でもある。スコットランドにおける行政記録情報のリンケージに関する組織的な連携において、Dibben教授が中心的な役割を果たしていることが伺える。

<sup>13</sup> 第4節で述べたように、イングランドにおいては、ONSがデータリンケージを行うだけでなく、公的統計の個票データや行政記録情報の提供業務を担っていることから、ONSがTREの役割を担っている。なお、下記のTREに関するADR UKのウェブサイトも参照されたい。

<https://www.adruk.org/data-access/trusted-research-environments/>

活用にあたっては、一般データ保護規則(GDPR)とデータ保護法(Data Protection Act)が法的根拠になっている。

つぎに、スコットランドにおけるデータリンケージの手続きについて見ていく。図 2 は、スコットランドにおけるリンケージのイメージ図を示したものである。図2によれば、最初にデータの供与者Aが、個人の識別子を Indexing team に提供すると、Indexing team は、個人に関する識別子群に関する確率的なマッチングを行い、仮名化を施すことによって、Anonymised ID を生成する。この ID は、直接識別子群から置き換えられたリンケージ用の ID である。つぎに、Indexing team によって、ID がデータの供与者に戻されると、データの供与者Aは、Anonymised ID を個票データ(図2では payload data)に付与し、個人の識別子を個票データから削除するという匿名化処理を行った上で、eDRIS team に送付する。eDRIS team は、Anonymised ID を用いて、複数のデータ間のリンケージを行う。研究者は、このリンケージされた個票データをセキュリティの管理がなされた環境でアクセスすることが可能になっている。

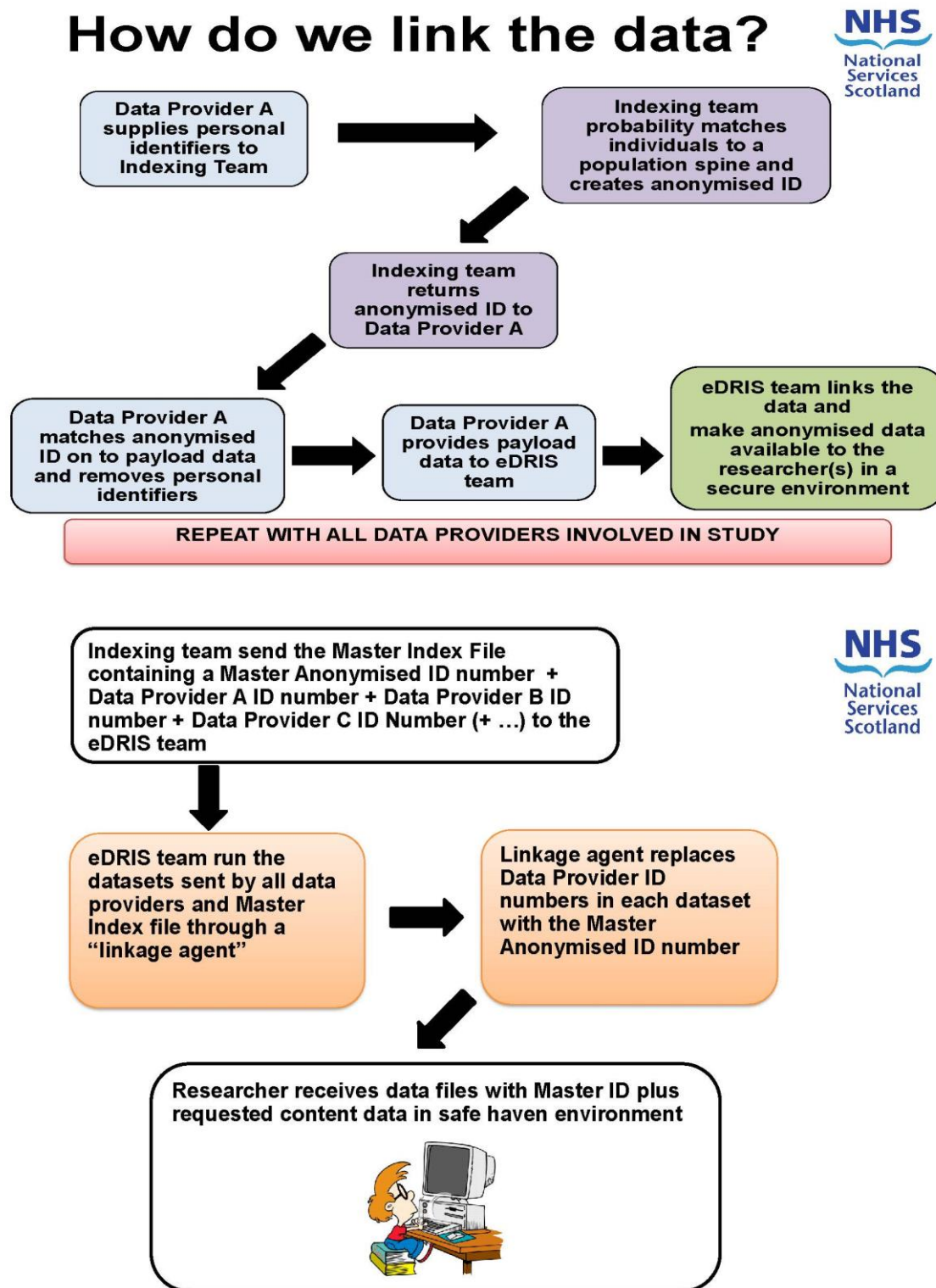
複数のデータ供与者からの個票データをリンケージする場合には、データ供与者の各個票データに対して付与したAnonymised ID の対応付けのために、Master Anonymised ID と呼ばれる基盤となる ID と各々の Anonymised ID の対応表 (Master Index File) を作成し、eDRIS team に送付する。eDRIS team は、すべてのデータ供与者から送付された Anonymised ID 付きのデータとlinkage agent を通じて提供されたMaster Index File を用いてリンケージを行う。このとき、linkage agent は、各データセットに含まれるデータ供与者向けの Anonymised ID を削除し、別の ID であるMaster Anonymised ID に置き換える。研究者は、この ID が付与された、研究者が必要としている変数を含む個票データをオンサイト施設やリモートアクセスのような安全な環境の下で利用することが可能になっている<sup>14</sup>。

つぎに、医療健康データと行政記録情報とのリンケージを行う場合という具体的なケースに焦点を当てて、データのリンケージの方法について概説する。図 3-1 から図 3-3 は、NHS が保有する医療健康データと行政記録情報等の NHS 管轄外の情報とリンケージを行うという具体的なケースを図示している。リンケージの作業は Indexer と Linker と呼ばれる役割を担う 2つの組織によって実施される。Indexer は、先述の NRS 内にある Index Team であり、Linker はPHS 内部のeDRIS 内に位置している Linkage Agent という部署が担っている。

最初に、Indexerの役割は、Linkerが個人を特定することなくリンケージ可能にするために、リンケージすべき対象を特定するための識別子の作成を行うことである。図 3-1 では NHS Scotland、警察(Police)、社会事業部(Social Service)の3つの公的機関からのデータが供与される例を表示している。図 3-1 で示されているように、データを提供する複数の機関から、リンケージされる予定のレコードに関連した、氏名や生年月日、住所といった直接的な識別子 (direct identifier、以下「DI」と呼称)が提供される(図 3-1①)。その上で、確率的マッチング (Probability Matching)等のマッチング技法をもとにDIによるリンケージが行われる(図3-1②)。マッチングされたレコードに関しては、Indexer によって、それぞれの DI に対応するリンケージ用の ID が付与される (図 3-1③)。図 3-1 において示されるプロセスで、リンケージの対

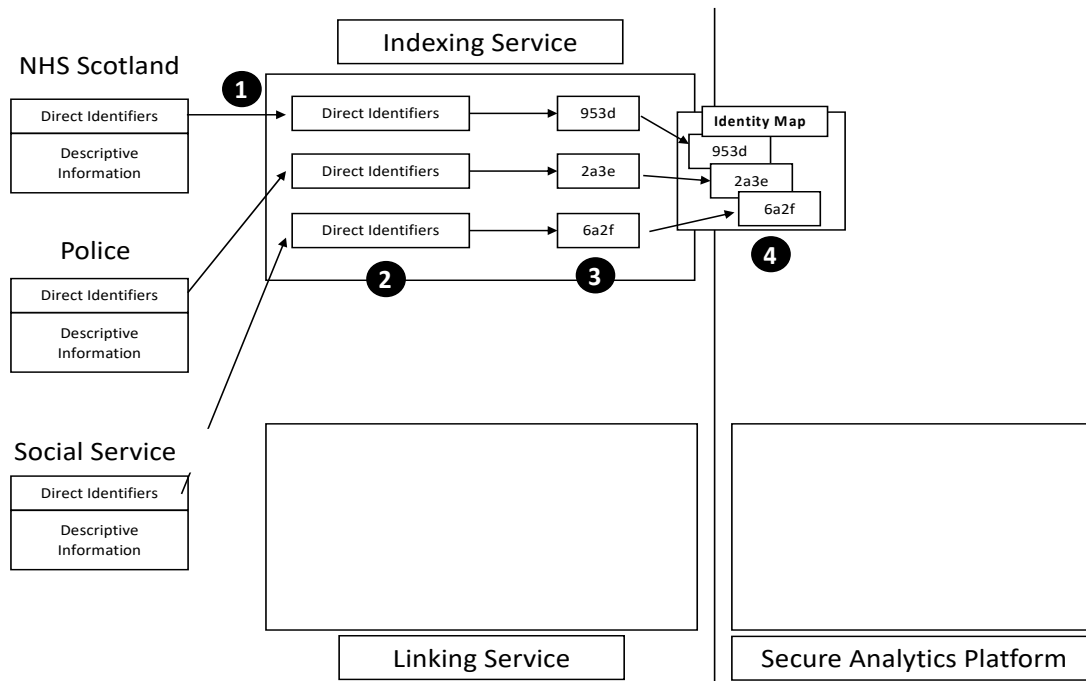
<sup>14</sup> TRE は、承認(認可)された研究者が安全な実行環境の中で実証分析を行うフレームワークを提示している。そして、個票データにリンケージ用の ID (Anonymised ID)とはさらに別の ID (Master Anonymised ID)が付与されることによって、対応表を利用者が入手することができず、利用者による再識別が困難になるような措置が採られている。そのため、利用の対象となるデータに対して仮名加工の処理のみが施されていても、追加的な匿名化技法の適用を必要としない程度にはリスクを軽減することが可能になっている。このことから、行政記録情報や医療健康データに対して実際に適用されている手法は「仮名化」のみであったとしても、実質的には「匿名化」が行われたデータが利用者に提供されるという解釈が成立しうる。

図2 スコットランドにおけるデータリンケージのイメージ図



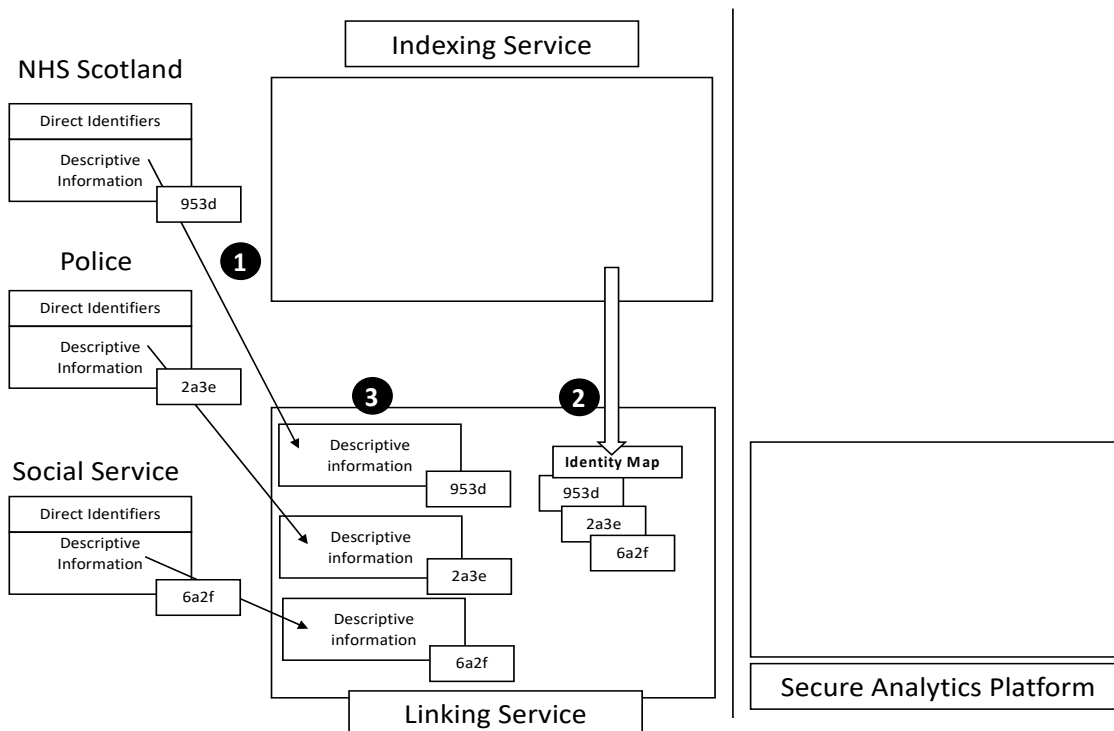
注 本図は、Public Health Scotland の Jackie Caldwell 氏から提供された資料に基づいている。

図 3-1 National Safe Haven で実施されるリンケージのプロセスにおける Indexer の役割



注 本図は、University of Dundee の Emily Jefferson 教授へのインタビュー調査(2022年9月15日)で入手した資料に基づいている。

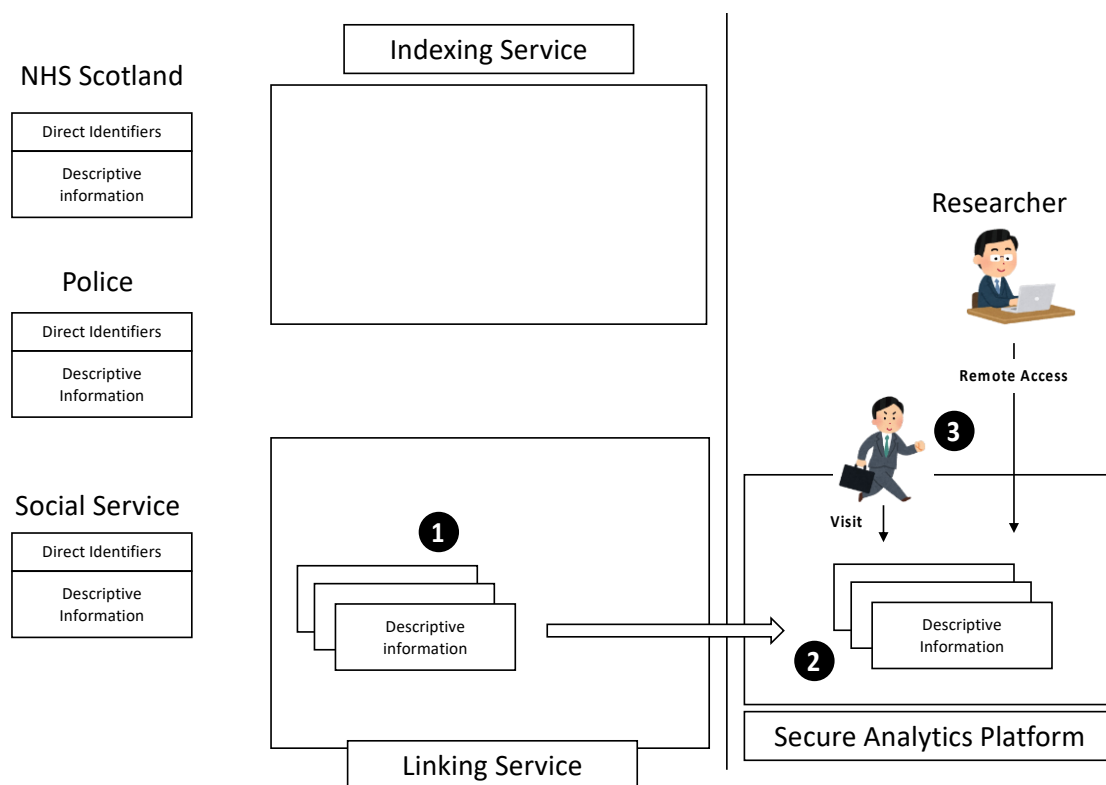
図 3-2 National Safe Haven におけるリンケージのプロセスにおける Linker の役割



注 本図は、University of Dundee の Emily Jefferson 教授へのインタビュー調査(2022年9月15日)で入手した資料に基づいている。



図 3-3 National Safe Haven における Linker によるリンケージデータの安全な分析環境への提供



注 本図は、University of Dundee の Emily Jefferson 教授へのインタビュー調査(2022年9月15日) で入手した資料に基づいている。

象となる個人において、NHS から提供されるデータの DI に対して'953d', Police から提供されるデータの DI に対して'2a3e', Social Service から提供されるデータの DI に対して'6a2f' というリンケージ用の識別子が、それぞれ付与されている。また、各 DI に対応した識別子が同一の個人に紐付けられているものであることが確認可能なように、Identity Map に DI の対応表が作成されている (図 3-1④)。データ提供元に対しては、それぞれの DI に対応したリンケージ用の識別子と対応する DI が戻される。この作業において Indexer は、DI 以外のデータにアクセスことはできない。

つぎに、Linker の役割は、個人を直接特定できる情報を用いることなく、リンケージを行うことである(図 3-2)。データの提供機関はそれぞれ、保有するデータから DI を削除するだけでなく、Indexer より提供されたリンケージ用の識別子を付与した上で Linker へ移送する (図 3-2①)。また Indexer から Linker に Identity Map(DI の対応表)が送付される(図 3-2②)。Linker は同一人物に属するリンケージ用識別子群について、Identity Map を用いて当該リンケージ識別子を紐付けた上で、対応する各個票データ間のリンケージを行う(図 3-2③)。

最後に、Linker によるリンケージデータの安全な分析環境への提供がなされる(図 3-3)。リンケージしたデータからリンケージ用の識別子を削除した個票データが(図 3-3①)、セキュリティの管理がなされた環境において利用可能なデータとして研究者に提供される(図 3-3②)。それによって、研究者はリモートで個票データにアクセスすることができる(図 3-3③)。

こうしたプロセスの特徴は、DI とそれ以外の個人に関する属性情報が直接連結される状態を回避するために、Indexer が DI 以外の個人に関する属性情報にアクセスできないだけな

く、Linker が個人を直接特定できる識別子にアクセスすることなく、リンケージ用の識別子の対応表を用いて、行政記録情報や医療健康データとのリンケージを行っていることである。こうしたプロセスを経ることによって、スコットランドにおいては、研究者が、分析対象となるデータに含まれる個人のプライバシーの保護を担保しつつ、行政記録情報と医療健康データに関するリンケージデータのアクセスを行うことができる。

このようにスコットランドでは、イングランドとは異なり、医療健康データと公的統計や行政記録情報といった社会経済に関するマイクロデータとのリンケージが可能な仕組みが、制度的組織的に整備されてきた。また、これらのデータのリンケージを行うための技術的な対応も、Indexer と Linker を別々の行政機関が担うことによって可能になっている。このように、Indexer と Linker のそれぞれについて異なる行政機関が業務を行うことで、Indexer は、個票データに含まれる属性情報を見ることなく、DI とリンケージ用の識別子の対応表に関する管理業務に専念することができる。その一方で、Linker は、DI を閲覧できないため、Linker によって個体が識別されるリスクは大きく低減される。その上で、リンケージ用識別子群を用いてデータ提供機関によって送付されたデータのリンケージを行うことが可能である。こうしたデータ提供機関が保有する個票データの安全管理の観点から、この仕組みにおけるメリットは大きいと考えられる。

先述のとおり、デンマーク等の北欧諸国においては、学術研究目的のために、医療健康データを含む各種の行政記録情報が共通の ID でリンケージされた上で、研究者に分析環境へのリモートアクセスの形で、個票データが広範に提供されている(木村他(2020)、伊藤(2020b))。スコットランドでは、北欧諸国のように各種の行政記録を保有し、様々な行政記録情報のリンケージを可能にするシステムは備わっていない。しかしながら、スコットランドにおいては、PHS、NRS および ADR Scotland が連携しながら、仮名化された ID を生成し、その ID でデータリンケージを行うことによって、NSH が備える信頼された研究環境の下で、北欧諸国と同様な医療健康データと社会経済データのリンケージデータに関する研究が可能になっている。

## 5. まとめ

本稿では、イギリス、特にスコットランドを例に、学術研究目的のための公的統計の二次利用や行政記録情報の利活用の現状を明らかにした。また、行政記録情報を対象にしたリンケージの追求とリンケージを対象にした実証研究が、イギリスにおいて広範に展開されていることを確認した。特に、スコットランドでは、医療健康データと人口社会、社会経済に関する行政記録情報のリンケージを可能にするためのセキュリティの管理がなされた環境の整備とリンケージを可能にするための ID の生成、および個人情報の秘密保護を確保した上で、ID とデータの伝送に関する技術的・組織的対応について論じた。

わが国では、事業所・企業系の公的統計や行政記録情報に関しては、企業の名称や住所、事業所母集団データベース等で把握可能な事業所番号、さらには法人番号といった識別子を用いてリンケージを行うことが可能であり、それらのリンケージデータの精度検証やリンケージデータを用いた実証研究も広範に行われてきた。しかしながら、国勢調査のような世帯・人口系のデータの場合、公的統計の個票データと行政記録情報の間でデータリンケージを行おうとしても、共通 ID 等の個体に関する識別子がリンケージのために利用できないことから、識別子を通じたリンケージは極めて困難なのが現状である。

個人に関するリンケージのための共通 ID が存在しないわが国の状況において、異種のデータのリンケージに関しては統計的マッチングを適用することは、実証研究においては現実

的な選択肢の1つになるだろう。その一方で、第1節で述べた、わが国における公的統計の二次的利用をめぐって2023年6月に出された「答申」には、「調査票情報に含まれる回答者ごとの連結(「紐付け」)を可能にするための方策」についての明記がなされている。学術研究目的のために、公的統計の調査票情報に関するリンケージを可能にする「紐付け」に関する具体的な議論が望まれる。さらには、本稿で議論した共通IDの生成に関するイギリスの事例を参考にしながら、わが国においても行政記録情報に関するリンケージを可能にする方策を模索する方向があってもよいと考える。

ただし、それぞれの行政記録情報にある直接的な識別子の仮名化によってリンケージのためのIDを生成し、信頼できる第三者機関における安全管理措置が取られた環境の下でデータリンケージを行う場合、わが国においてその実用化を目指そうとすれば、検討すべき課題は少なくないと思われる。例えば、本稿で議論したように、スコットランドにおいて、Indexerとして機能するNRS内部に設置されたIndex Teamや、Linkerとしての役割を担うeDRIS内に位置するlinkage agentのような機関に関して、どのような法的根拠に基づいてわが国で設置することが可能かについての議論が必要になるだろう。わが国における大規模データの利活用のさらなる展開を図るためには、こうした点も含め、中・長期的な観点から今後さらなる検討が求められる。

## 謝辞

本稿の作成にあたって、イギリス国家統計局(ONS)(2023年6月9日)、西イングランド大学ブリストル校(UWE Bristol)(2023年6月6日)、およびエディンバラ大学(2023年6月15日)においてヒアリング調査を実施した。インタビューに応じていただいた、Bill South氏(ONS)、Iain Dove氏(ONS)、Hayley Holgate氏(ONS)、Felix Ritchie教授(UWE Bristol)、Jim Smith教授(UWE Bristol)、Chris Dibben教授(エディンバラ大学)、Jackie Caldwell氏(Public Health Scotland)、Simon Rogers氏(NHS National Services Scotland)、Gillian Raab博士(エディンバラ大学)にお礼を申し上げたい。また、National Records of Scotland(NRS)のLiam Cavin博士からは、NRSの活動状況に関するメールでの質問に対して、丁寧な回答をいただいた(2023年6月21日付)。記して謝意を表したい。

## 参考文献

- [1] 伊藤伸介(2011),「わが国におけるマイクロデータの新たな展開可能性について—イギリスにおける地域分析用マイクロデータを例に—」, 明海大学『経済学論集』Vol.23, No.3, pp.36-54.
- [2] 伊藤伸介(2016a),「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—」, 『経済学論纂(中央大学)』第56巻第5・6合併号, pp.1-19.
- [3] 伊藤伸介(2016b),「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, pp.233-249.
- [4] 伊藤伸介(2018),「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, pp.77-101.
- [5] 伊藤伸介(2020a),「諸外国における公的統計と行政記録情報の二次利用に関する展開方向」, 『経済学論纂(中央大学)』第61巻第2号, pp.1-16.
- [6] 伊藤伸介(2020b),「デンマークとオランダにおける医療健康データの二次利用について」, 『日本統計学会誌』, 第50巻第1号, pp. 109-138.

- [7] 伊藤伸介(2022),「マイクロデータの匿名化と統計情報の秘匿可能性について」『経済学論纂(中央大学)』第63巻1・2合併号, pp.1-23.
- [8] 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, pp.187-208.
- [9] 伊藤伸介・寺田雅之(2023),「海外における公的統計に関するプライバシー保護の現状—アメリカとイギリスの事例をもとに—」『統計研究彙報』第80号, pp.117-136.
- [10] 木村映善・大寺祥佑・佐々木香織・黒田知宏(2020),「フィンランドにおける医療分野レジスタとデータ提供の状況」『日本統計学会誌』, 第50巻第1号, pp. 47-80.
- [11] 佐々木香織・大寺祥佑・木村映善(2020),「より包括的で正確な医療統計を可能とする社会・制度基盤に向けた一考察—イギリスの England における医療情報二次利用に関する調査・事例研究から—」『日本統計学会誌』, 第50巻第1号, pp. 81-108.
- [12] Bujnowska, A., Espelage, F., Stocchi, M.(2023), “Remote access to European microdata”, Paper presented at UNECE Expert Meeting on Statistical Data Confidentiality 2023, pp.1-6.
- [13] Brenzel, H., Cramer, K., Güttgemanns, V., Mathes, M. (2023), “Remote Access for Scientific Use Files – a New Pathway for German Official Statistics Microdata Access”, Paper presented at UNECE Expert Meeting on Statistical Data Confidentiality 2023, pp.1-6.
- [14] Cranswick, K. (2019), “Virtual data labs - A more flexible approach to access Statistics Canada microdata”, Paper presented at Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, pp.1-13.
- [15] Desai, T., Ritchie, F. and Welpton, R. (2016), Five Safes: Designing Data Access for Research, Economics Working Paper Series 1601, University of the West of England.
- [16] 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.