

イギリスの公的機関による 大規模リンケージデータの作成の現状

2025年4月1日

中央大学経済学部

伊藤 伸介

1.はじめに

- わが国における事業所・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たリンケージを行うことが可能である。

→個票データのリンケージによって作成されたデータ(以下「リンケージデータ」)については、その精度の検証を含む実証研究が広範に展開されてきた(例えば、村田・伊藤(2016)等)。

- 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同士、あるいは公的統計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、リンケージ用の共通ID等の個体識別子が利用できないことから、識別子を用いたリンケージは困難である。

- わが国と同様に、共通IDについての仕組みが存在しないイギリスでは、2018年に施行された**Digital Economy Act(DEA)**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面で整備された。
- イギリスの場合、資金提供を担う研究インフラ支援機関(funding agency)である**ADR UK**の支援の下で、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展開されている(伊藤(2020a))。
- イギリスでは、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、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))。
- 本報告では、イギリスを例に、海外の公的機関における大規模リンケージデータの作成に関する最近の動向について明らかにする。

2. 海外の公的大規模データにおけるデータリンケージサービスの状況

- 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), 伊藤(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。

世帯・人口系の公的大規模データにおけるリンケージの状況(伊藤他(2024), 伊藤(2024), 伊藤(2025))

①個人に付与されている共通ID(個体識別番号)をもとに、統計作成部局による公的統計や行政記録情報のリンケージが可能な国々

フィンランド、**デンマーク**、スウェーデン、オランダ等

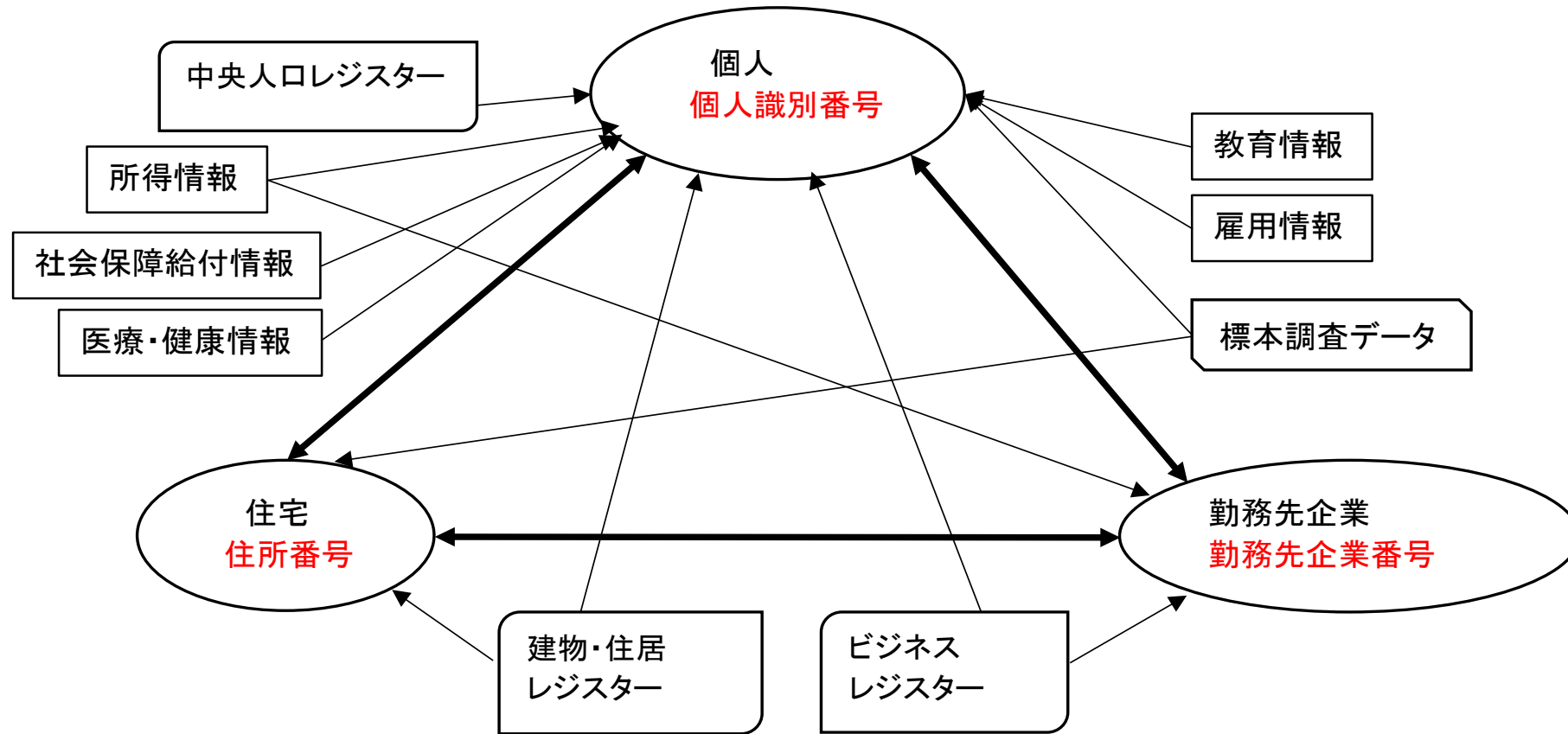
②リンケージ用のIDを生成することによって、統計作成部局による公的統計や行政記録情報のリンケージが可能な国々

イギリス、アメリカ、カナダ、オーストラリア、ニュージーランド等

③リンケージ用のIDが存在せず、統計作成部局やデータ提供機関がデータリンケージを行わない国々

ドイツ、フランス等

図1 デンマークにおけるデータリンケージの状況



出所 伊藤(2020b)、図2

※人口社会情報(結婚と離婚、出生と死亡、移住等)、労働・所得情報(就業、職業訓練、賃金、家族所得等)、教育に関する情報(最終学歴、教育歴等)、企業経済情報(企業の雇用、購入・販売等)、医療健康情報(国民患者レジスター(National Register of Patients)、医薬品統計レジスター(Register of Medical Product Statistics)等)といった様々な情報を対象に、**仮名化されたID**にしたがって、デンマーク統計局がリンケージを行う。研究者は、リンケージされたデータにリモートでアクセスすることが可能である、

3.イギリスにおける公的大規模データにおけるリンケージの動向

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が展開(伊藤(2016a), 伊藤他(2024), 伊藤(2025))されてきた。

2011年:社会経済研究協議会(Economic and Social Research Council=ESRC)の主導によって、[Administrative Data Taskforce\(ADT\)](#)が設置

→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

2012年12月: ADTが報告書『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)が刊行

→「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

2014年10月: [Administrative Data Research Network\(ADRN\)](#)が創設され、行政記録情報のリンケージに関する提供サービスが開始

→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内部のオンライン施設でリンケージされた行政記録情報の非識別データ(deidentified data)を利用することができた。

2018年: [ADRN プロジェクト](#)は終了

ADR UKプロジェクト

- イギリス全体(UK)を統括する **ADR UK** 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を展開している(伊藤(2020a))。

⇒イングランドの場合、**イギリス国家統計局(Office for National Statistics=ONS)**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非識別データの提供を一元的に担っている。

- 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。

Digital Economy Act 2017について

・イギリスでは、2018年にDigital Economy Act(以下DEA)が施行



研究目的のために、あらゆる統計データ、行政記録データ、政府によって(潜在的に)生み出されるビッグデータの利用(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(民間企業が所有するパーソナルデータについては対象外)

第5章 研究目的のための共有

第64条 研究目的のための情報の開示(disclosure)

第1項 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。

(以下省略)

※Digital Economy Act(DEA)に基づいて行政記録情報を利用するにあたっては、**認可された研究者(accredited researcher)** になるための資格取得が求められる。

⇒ONSとしては、利用者が「承認された研究者(approved researcher、イギリス統計法における概念)」よりも「認可された研究者」の資格を取得することが通常考え方となっている。

・行政記録情報の大部分については、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。

→DEAの範囲外である医療データでもSRSで提供可能なものがある。

ONSによる公的大規模データのリンケージの手順

①ONS内部で個票データの提供を担うSRS(=Secure Research Service)とは異なる**リンケージ担当部署**において、リンケージの対象となる複数のマイクロデータから直接的な個体識別子(Direct Identifier=DI)のみを切り離す。

←Data Access Platform(DAP)の中に**データリンケージチーム**がある。DAPには、データアナリスト等もいるが、データリンケージチームに関しては20人～30人のチームで構成されている。

②リンケージチームは、仮名化担当チーム(indexing team)とマッチング担当チーム(matching team)に分かれている。仮名化担当チームがDIに対して仮名化を行うことで、そこからリンケージ用の仮名化IDを生成する。この仮名化IDの生成に関する作業においては、DIを用いてマッチングを行った後に、リンケージ用の仮名化IDの対応表のみが保存される。

③マッチング担当チームが、リンケージ用の仮名化IDの対応表に基づいて、複数の公的大規模データ間のリンケージを行う。

④リンケージが行われた後に、リンケージ用IDが削除された非識別データがSRSに保管される。DEAIに基づき、認可された研究者(accredited researcher)になるための資格取得の手続きを行った上で、研究者は、SRSを通じてこのリンケージデータにアクセスすることができる。したがって、SRSでは、非識別データの提供業務を担っているのみであり、データリンケージに関する業務には携わっていない

※一度リンケージされたデータは、他の研究グループも申請することで、利用することが可能になっている。

- ・リンケージの作業において重要な点は、仮名化の段階で、最初からDI情報と属性情報が切り離されていることである。

⇒DIを仮名化する作業と個々のデータに仮名化IDを付与する作業についてはデータリンケージチームの別々の職員が担当している。

- ・マイクロデータ間のマッチングにあたって、DI情報は、DAPでDI情報の仮名化に直接携わっている職員以外は、閲覧できないようになっている。

←仮名化の後、DI情報はONS内部で消去されていることに留意する必要がある。

- ・リンケージに関しては、「オーダーメイドのリンケージ(bespoke linkage)」を行っていることが特徴的である。

⇒データリンケージの作業を行う上で、リンケージの対象となる個票データの数が増えるほど、時間を要することを意味している。

・マッチングの手法については、データの特性に依存するが、データリンケージの担当業務を担っているONSの担当者によれば、最初に**確定的なリンケージ(deterministic linkage)**を行うことを検討し、つぎに、確定的なリンケージが困難な場合に、**確率的なリンケージ(probabilistic linkage)**を実行するとのことである。

・ONSは、イングランドとウェールズにおける2022年退役軍人調査(Veterans' Survey 2022)の代表性とカバーレッジの程度を推定するために、2022年退役軍人調査と2021年イギリス人口センサスを対象に、2つの統計調査の個票レコード間で**確定的なリンケージ**を実施した上で、リンケージの精度評価を行っている。

→リンクがなされていないレコード群については、精度検証を目的とした確率的なリンケージが実行された。

スコットランドにおける行政記録情報の提供の特徴(伊藤他(2024))

- ・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、行政記録情報と国民健康保険サービス(National Health Service＝NHS)が保有する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。

- ・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。

⇒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。

- ・**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。

- ・**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

←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について

- TREは、**5つの安全モデル(Five Safes Model)**に基づいて設計されている。
 - ⇒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)**データリンク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。
- TREについては、安全性が確保された施設・分析環境(Secure Research Environment)と捉えることもできるが、広義としては5つの安全モデルに基づいて行われるマイクロデータ提供のための**フレームワーク**と考えられる。
- イングランドにおいては、ONSがデータリンク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。
 - SRSの担当者によれば、SRSはTREであり、TREとは安全にデータを提供する環境を研究者に提示することを意味している。
 - ⇒オンサイト施設やリモートアクセスはデータを提供する環境に該当するが、**安全な利用者や安全なプロジェクト、安全なデータと安全な分析結果がセットになって、TREを準備することが**できる。

参考 5つの安全モデルについて(Five Safes model)(Ritchie(2008), Desai et al.(2016),伊藤(2016b), 伊藤(2018), 伊藤他(2024))

* 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

①安全なプロジェクト(safe projects)

→妥当な統計目的のために個票データのアクセスが行われること。

②安全な利用者(safe people)

→研究者は個票データを適切に利用し、利用手続きに従うことについて、信頼されていること。

③安全なデータ(safe data)

→データそれ自体が、個人情報が見えないデータであること。

④安全な施設(safe settings)

→個票データのアクセスに関する技術的な管理措置によって、容認されていないデータの移動が回避可能なこと。

⑤安全な分析結果(safe outputs)

→統計分析の結果に個人情報が見えるような結果が含まれないこと。

※分析結果に関する最終成果物については、担当者がチェックを行った上で公表することが可能



安全な利用(safe use)

5つの安全モデルの考え方によれば、ひとつ1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。

⇒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。

- 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: **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)
⇒ **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(trusted third 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。
- 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。
- 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。
⇒ 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。

図2-1 スコットランドにおけるデータリンケージのイメージ

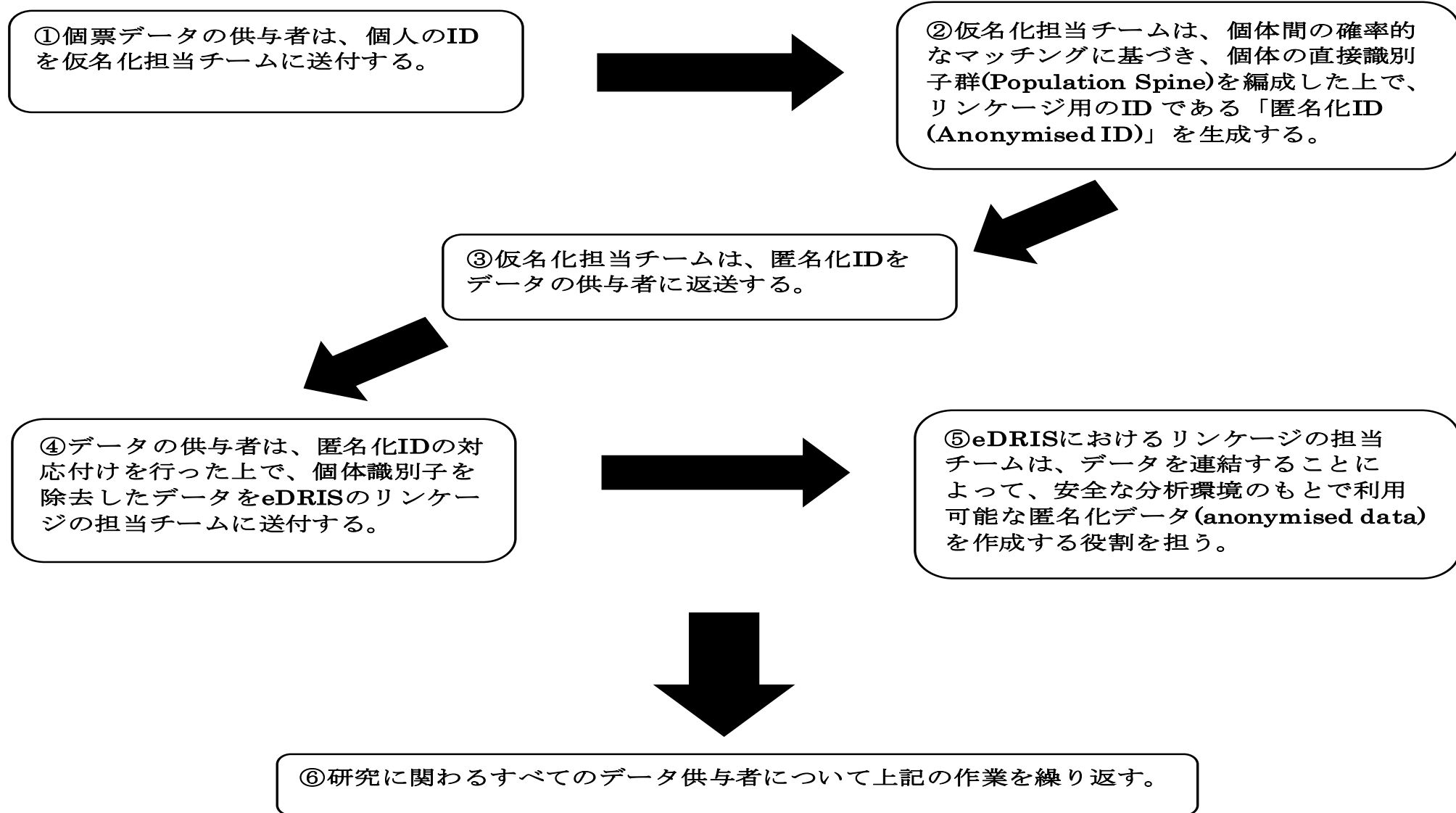
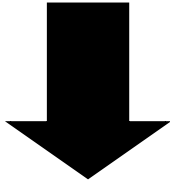


図2-2 スコットランドにおけるデータリンケージのイメージ(続)

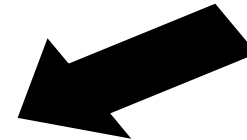
⑦仮名化担当チームは、共通匿名化ID(Master Anonymised ID)、およびそれぞれの供与者が提供する、データに付与された仮名化IDが組み合わされた「匿名化IDの対応表(Master Index File)」をeDRISに送付する。



⑧eDRISにおいては、linkage agentを通じて、すべてのデータ供与者が送付したデータセットと匿名化IDの対応表を用いたデータリンケージが実行される。



⑨linkage agentが、各データセットに付与された仮名化IDの共通匿名化IDへの置き換えを行っている。



⑩研究者は、安全な分析環境のもとで、申請した共通匿名化IDが付与されたリンケージ済みの匿名化データにアクセスすることができる。

データリンケージの手順(1)

- ①データの供与者が、データに含まれる個別主体の直接識別子を**仮名化担当チーム(Indexing team、NRS内部に設置)**に提供すると、仮名化担当チームは、**population spine**と呼ばれる個別主体の直接識別子群を対象に、個体間の確率的なマッチングを実行する。つぎに、仮名化を行うことによって、リンケージ用のID としての「**匿名化 ID(Anonymised ID)**」を生成する。
- ②仮名化担当チームによって、匿名化IDがデータの供与者に戻されると、データの供与者は、匿名化IDを個票データに付与し、直接的な個体識別子が個票データから削除したデータを**eDRIS**に送付する。なお、eDRISにおける**リンケージの担当チーム**は、匿名化IDを用いてデータリンケージを行うことによって、リンケージされた匿名化データを編成する役割を担う。

データリンケージの手順(2)

③個票データをリンケージする場合には、各データ供与者から送付された個票データに付与した匿名化IDのそれぞれを対応付けるために、**仮名化担当チーム**が「**共通匿名化ID(Master Anonymised ID)**」と各々の匿名化IDの**対応表(Master Index File)**を作成した上で、eDRISに送付する。

④eDRISにおいては、**linkage agent**を通じて、すべてのデータ供与者から送付された匿名化ID付きのデータセットと匿名化IDの対応表を用いてリンケージが行われる。各データセットに含まれるデータ供与者向けの匿名化IDはlinkage agentによって削除され、共通匿名化IDに置換される。

⑤研究者は、共通匿名化IDが付与されたリンケージ済みの匿名化データ(anonymised data)をオンサイト施設やリモートアクセスのような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。

・スコットランドにおいて公的大規模データおよび医療健康データのリンケージを行う上で、TREは、認可された研究者が安全な分析環境において実証研究を行うためのフレームワークを提示している。

・研究者がアクセス可能な個票データには匿名化IDとは異なる共通匿名化IDが付与され、利用者による再識別が困難な加工が施されている。

→研究者にとって利用の対象となる個票データに仮名化が適用されており、さらなる匿名化処理が求められないレベルにまで再識別のリスクが低減されたものとみなされる。

⇒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」のみであるが、研究者にとっては「匿名性」を有する非識別データが、研究者によって利用可能になっている。

4. イギリスにおけるIntegrated Data Serviceについて

- イギリスにおけるデータリンケージの新たな展開として注目されるのが、イギリス政府主導で進められている**Integrated Data Service(IDS)**と呼ばれる国家プロジェクトである。
- IDSでは、公的大規模データだけでなく、医療健康データや社会調査の個票データも対象にする形で、各種のデータのリンケージとリンケージデータを用いた分析環境を提示する。
- IDSにおいて、**ONS**は各種の行政記録情報や公的統計ミクロデータを対象にしたデータリンケージの中核的な役割を担っている。

IDSが想定している利用者は、2024年2月時点では、大学や研究機関に属する研究者ではなく、政府機関に所属するデータアナリストである。

←IDSでは、政策立案、政策評価といった利用目的に適合したリンケージデータを作成・提供することを指向している。

・IDSでは、ONSが、①各種の個票データの中のDI情報を仮名化されたIDに変換するという匿名加工を行い(anonymised)、DI情報との対応表に基づいて、リンケージ用の仮名化IDに置き換え(indexed)、②安全性が確保された環境の下で共通の仮名化IDに基づくリンケージを行う(securely linked)。

⇒リモートアクセスのような安全な分析環境の下で、リンケージデータを用いた実証分析が可能になる。その意味では、現行のSRSの仕組みをそのまま引き継いでいる。

・リンケージデータを用いた分析結果のチェックをONSの担当者が行うことで、分析結果の外部への持ち出しが可能になるが、この審査プロセスについても、IDSはSRSでの業務を引き継いでいる。

→IDSも5つの安全モデルに対応した安全な分析環境に関するフレームワークを備えており、SRSと同様にTREだということができる。

- 法制度的な観点から見ると、IDSの運営は、行政記録情報のリンケージの促進を指向して施行されたDEAに基づいている。

- DEAの対象外となっているデータは、IDSで利用可能なデータとはならない。

- 2024年2月時点では、IDSの対象となるデータについては、医療健康データの多くは含まれておらず、スコットランドと北アイルランドの公的大規模データも該当しないことから、公的統計ミクロデータ、行政記録情報や医療健康データのすべてがIDSにおけるデータ統合の対象にはなっていない。

5. おわりに

- 本報告では、イギリスを例に、公的機関による大規模リンケージデータの作成の動向を明らかにした。
- イングランドとスコットランドの両方におけるデータリンケージの特徴は、個体識別子に関する情報と属性情報を切り離した上で、個体識別子から仮名化されたリンケージ用IDを生成する仮名化担当チームと、仮名化されたIDで属性情報を含むデータのリンケージを行うマッチング担当チームに業務が分けられていることである。それによって、前者のチームは属性情報を確認することは不可能であり、後者のチームは、個々の直接識別子と仮名化IDとの対応関係を把握することは困難になっている。
- イングランドではONS内部で、仮名化業務とマッチング業務の両方が担われているが、スコットランドにおいては、仮名化を担当するNRSと、仮名化されたIDに基づくデータリンケージの役割を担うeDRISに組織が分かれている。このような措置によって、個体情報の秘密保護を担保するための技術的・組織的な対応が可能になっている。

- イギリスで現在注目されているIDSプロジェクトは、イギリスで展開されているリンケージに基づくデータ統合の発展的な形態だと言える。
- IDSは、現在進行中の国家プロジェクトであることから、IDSの対象となるデータ群がどのような形で連結可能になり、大規模なリンケージデータとして展開されていくのか、今後の動向を注視していきたい。
- わが国においてイギリスの事例を参考にしようとすると、イギリスで明確に役割分担がなされている、個体識別子の仮名化IDの置換や仮名化IDの対応表の管理に関する業務、さらには仮名化IDに関するマッチング業務をどのような機関が担うかについて、統計法制度的かつ組織的な観点、さらには統計技術的な観点から、方法的な可能性についての検討が求められるだろう。

参考文献

- 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.
- 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.
- 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.
- 伊藤伸介(2020a)「諸外国における公的統計と行政記録データの二次利用に関する展開方向」『経済学論纂』第61巻第2号, pp.1-16.
- 伊藤伸介・木村映善・佐々木香織(2024)「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—」『統計研究彙報』第81号, pp.89-108.
- 伊藤伸介(2024)「海外におけるマイクロデータの利活用推進に向けた取り組み状況—リモートアクセスを中心に—」『経済学論纂(中央大学)』第65巻1号, pp.199-218.
- 伊藤伸介(2025)「海外における公的大規模データのリンケージの動向—イギリスを例に—」『経済学論纂(中央大学)』第65巻5・6合併号, pp.187-200.
- 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.
- 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.
- 村田磨理子・伊藤伸介(2016)「事業所・企業系のマイクロデータを用いたデータリンケージの可能性—賃金構造基本統計調査を例に—」,『統計学』110号, pp.1-17.
- 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.
- 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.