

海外における公的統計マイクロデータの 作成・提供の現状

2023年3月6日

中央大学経済学部

伊藤 伸介

1.はじめに—わが国における公的統計のマイクロデータの提供状況

・わが国では、2007年に統計法(平成19年法律第53号)が成立してから、公的統計の二次的利用の推進が図られてきた。

→わが国においても、社会経済の様々な分野において、数多くの実証的なマイクロデータ分析が行われている。

・『公的統計の整備に関する基本的な計画』(第Ⅲ期基本計画)(2018年3月、2020年6月一部変更)において、EBPM(=Evidence Based Policy Making、証拠に基づく政策立案)の推進が強く指向されていることから、公的統計のマイクロデータを用いた分析結果に基づく政策立案やマイクロデータを用いた政策評価についても、社会的な注目が高まってきた。

・現行の統計法(平成十九年法律第五十三号)に基づき、わが国では、公的統計のマイクロデータが提供されている。

①調査票情報(個票データ)の提供

(1)磁気媒体による調査票情報の提供(法第33条)

・実証分析を行う上で**最低限必要な調査事項(変数)**のみの利用を想定し、統計作成部局が調査票情報の利用目的との適合性の観点から、利用申出において実証分析で作成する集計表と分析に用いるモデルの内容を審査する**事前審査方式**

(2)オンサイト施設による調査票情報の提供(法第33条、法第33条の2)

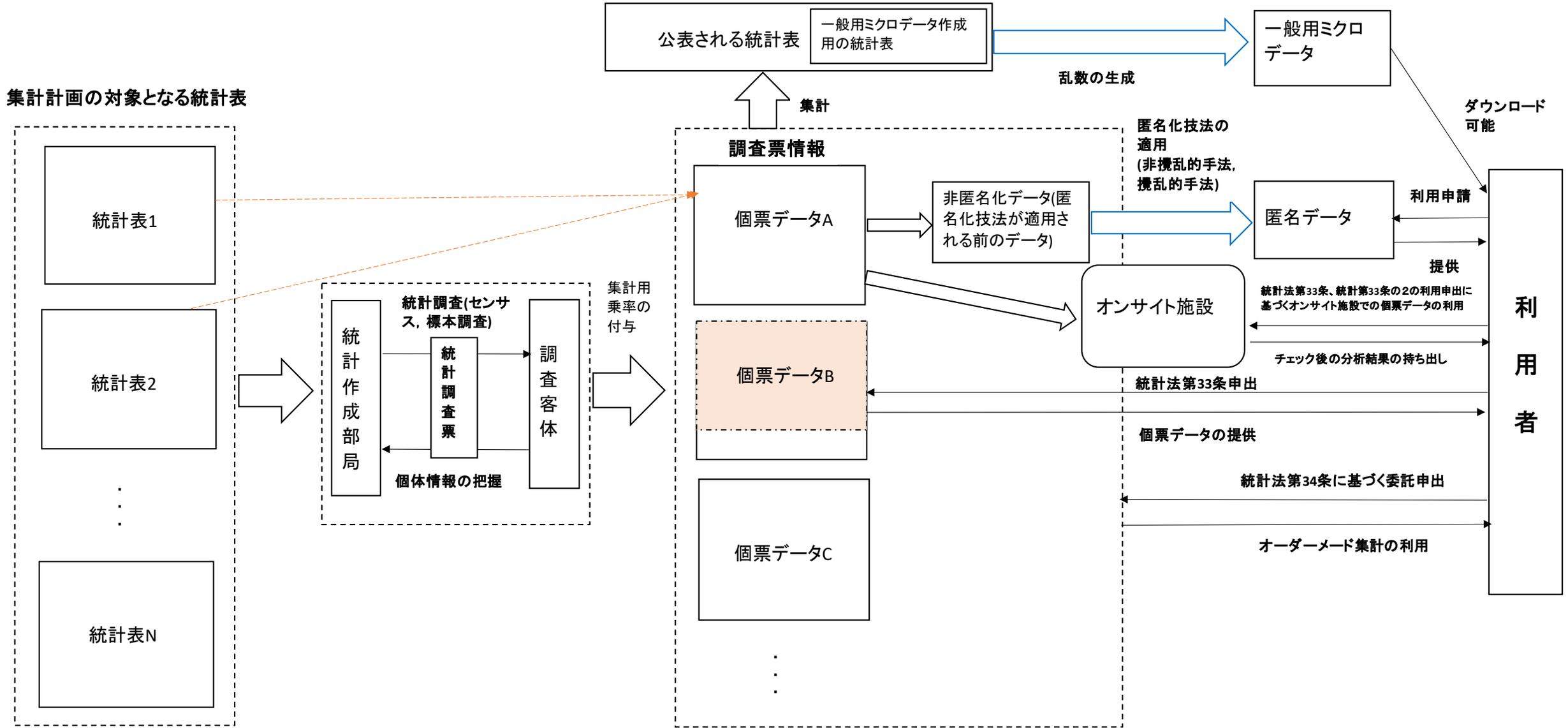
・オンサイト施設からの分析結果の持ち出しにおいて、分析結果についての安全性を確認するために、分析結果の秘匿性に関するチェックを行う**事後審査方式**

⇒世帯・人口系および事業所・企業系の統計調査のマイクロデータが利用可能

②匿名データの提供(法第36条)

③オーダーメイド集計(法34条)

図1 わが国の公的統計マイクロデータの作成・提供についての概略図



2. 欧米諸国における公的統計マイクロデータの作成・提供

・欧米諸国においても、様々なチャンネルを通じて、公的統計マイクロデータの作成・提供が行われてきた。

① 個票データ(非識別データ、deidentified data)へのアクセス

(1) Research Data Centre(RDC)におけるセキュアな環境(オンサイト施設)での利用

(2) リモートエグゼキューションにおける個票データの利用

- ・リモートアクセス(remote access)
- ・プログラム送付型リモートエグゼキューション
- ・オンデマンド集計(リモート集計)

② 匿名化措置が施された学術研究用の公的統計マイクロデータ(以下「匿名化マイクロデータ(anonymized microdata)」と呼称)の作成・提供

・学術研究用ファイル(Scientific Use File=SUF)の提供

⇒ 利用申請が必要であり、利用にあたってライセンス(誓約書)が必要

③ 一般公開型マイクロデータ(public use microdata)の作成・提供

・一般公開型ファイル(Public Use File=PUF)という形での提供

⇒ 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。

利用登録が必要な場合もあり

図2 公的統計の提供形態に関する概略図

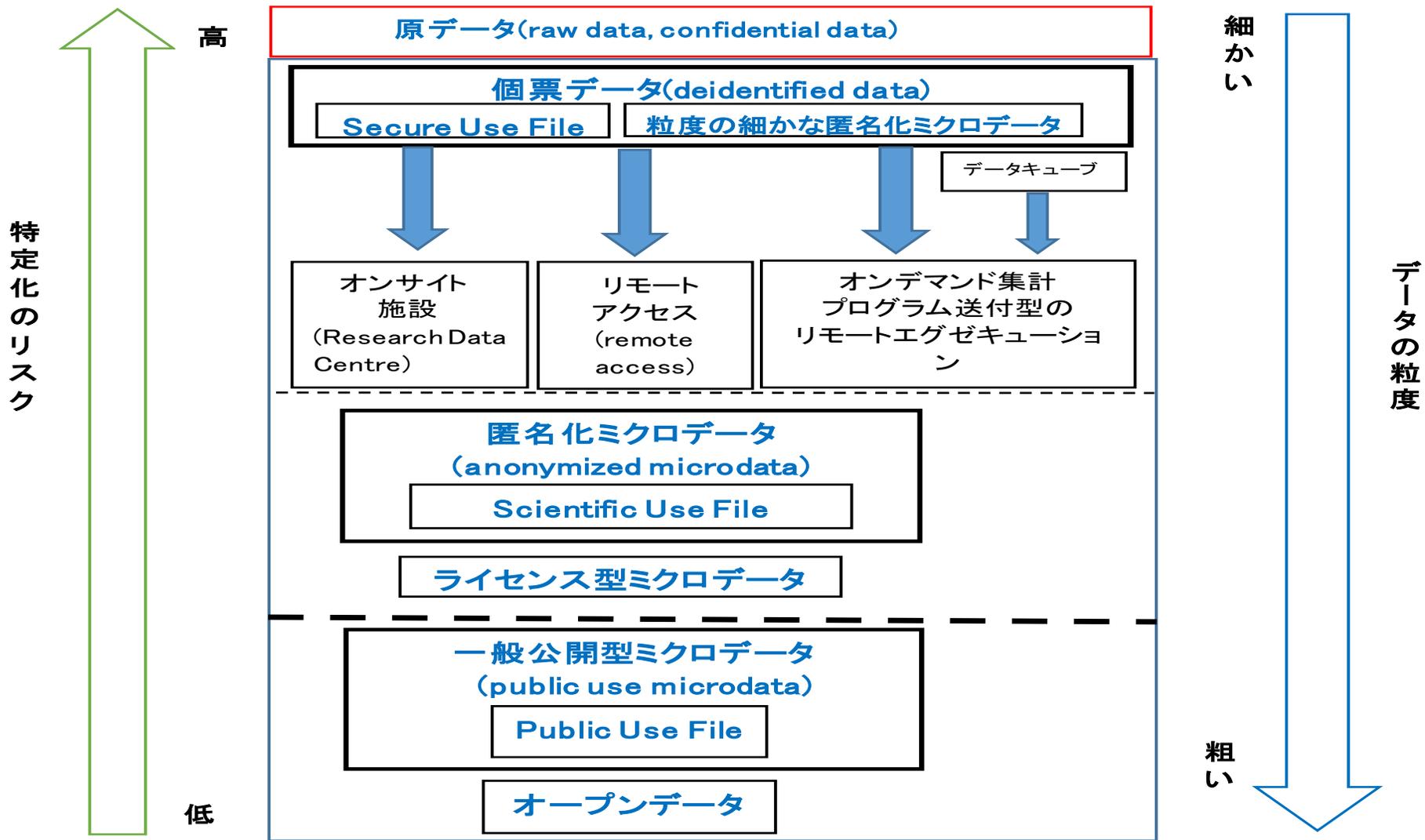


表1 海外の主な統計作成部局におけるマイクロデータの提供形態

データの種別/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：UKDSとはイギリスエセックス大学のU.K. 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(一部を除く))。

出所 伊藤伸介(2020)「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(中央大学)』第61巻第2号, 表1を一部修正

諸外国におけるマイクロデータ提供の類型化

① 個票データのアクセス、SUF、PUFといった複数のチャンネルによるマイクロデータの提供

⇒イギリス、ドイツ、オランダ等

② 個票データのアクセスと一般公開型マイクロデータの公開の2つのタイプによる提供

⇒アメリカ、カナダ

③ 個票データのリモートアクセスサービスのみでの提供

⇒デンマーク

※わが国の場合、調査票情報の提供、匿名データの提供と一般用マイクロデータの公開の3つのタイプの提供がなされているが、①に整理することができると思われる。

3. 公的統計マイクロデータの提供状況 ①イギリスの事例*

イギリスにおけるマイクロデータ提供の現状

1) 個票データの提供サービス

- **Secure Research Service(SRS)**において、オンサイト施設のようなセキュアな環境による個票データの提供

←UK Data Serviceのthe Secure Lab等による**リモートアクセスによる提供**

- **行政記録情報、公的統計マイクロデータとのリンケージされた行政記録情報、行政記録情報同士でリンケージされたデータの提供**

⇒2018年の**Digital Economy Act 2017**の施行と**ADR UK**による行政記録情報のリンケージサービス

2) 匿名化マイクロデータの提供

⇒U.K. Data Serviceにおけるライセンス(End User Licence)を必要とするタイプのマイクロデータ(「ライセンス型マイクロデータ」)の提供

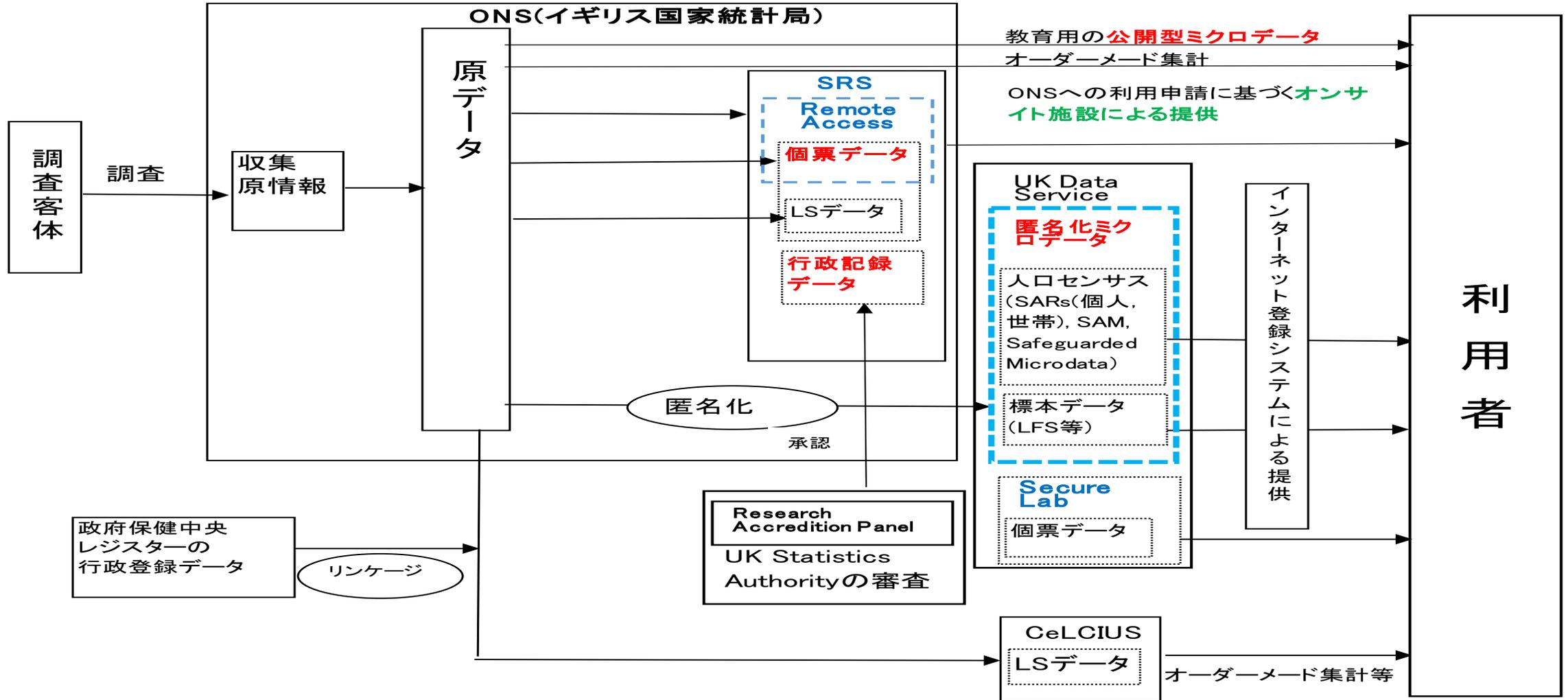
3) 一般公開型マイクロデータの公開

Ex. 人口センサスの教育用マイクロデータの作成・提供

* 伊藤伸介(2020)「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(中央大学)』第61巻第2号

伊藤伸介(2022)「ヨーロッパにおける大規模データのプライバシー保護の新たな動向」『統計』2022年8月号

図3 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, CeLCIUS: Centre for Longitudinal Study Information and User Support

出所 伊藤伸介(2020)「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(中央大学)』第61巻第2号, 図2を一部修正

参考 イギリス統計法(The Statistics and Registration Service Act 2007)

第39条 個人情報秘密保護

(中略)

(2)この条文における「**個人情報(personal information)**」とは、(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を意味する。

(中略)

(4)条文(1)は以下の開示には適用されない。

(中略)

(i)承認された研究者(approved researcher)によって行われる場合

(5)条文(4)(i)の目的に関して、「**承認された研究者**」とは、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。

(以下省略)

※第39条においては罰則規定も明記

参考 5つの安全モデル(Five Safes model)*

* 個票データのアクセスは、以下の5つの安全性に関する考え方に基づいて行われる。

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露見され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)

「承認された研究者」あるいは「認可された研究者(accredited researcher)」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。

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

- ・研究プロジェクトの目的や概要
- ・研究プロジェクトの期間
- ・モデル分析の手法
- ・利用する個票データの名称、年次
- ・使用するソフトウェア等

※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を検討されたうえで、容認された場合に個票データのアクセスが行われる。

* 縦断的なデータ構造を備えるLSデータ(ONS Longitudinal Study of England and Wales)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。

オンサイト施設における公的統計の個票データの提供サービス

・承認された研究者(approved researcher)あるいは認定された研究者(accredited researcher)の資格を取得し、申請している研究プロジェクトが、UK Statistics AuthorityのResearch Accreditation Panel(ONSに設定されていたMicrodata Release Panelの後継の審査委員会)による審査を通ると、ONSに設置されているオンサイト施設であるSafe Roomsやセキュアな環境としてのSafePodsにおいて個票データの利用が可能

→直接的な識別子を用いてONSがリンケージを行っ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←仮名化されたIDを用いて、研究者自身がSRS内でマッチングを行うことも可能。

・Assured Organisational Connectivity (AOC) agreementをONSとの間で締結した研究機関に所属する研究者は、所属する機関でリモートで個票データにアクセスすることが可能

・SRSでは、個票データに関するメタデータカタログのウェブページが整備されている。

⇒利用者が個票データに関する一括検索が可能

・利用者の分析結果については、秘匿性に関するチェック(output checking)が求められる。SRSの担当部局だけでなく、集計結果表やマイクロデータの秘匿に関する担当部局と協力しながら、マニュアルでチェックを行っている。

←公表前の分析結果(中間生成物)の持ち出し(pre-publication clearance)も可能であるが、これについては、研究プロジェクト内でのみ共有することが容認される。なお、プロジェクト終了後に破棄することが求められる。

参考 Digital Economy Act 2017について

- ・イギリスでは、2017年4月にDigital Economy Act 2017の法案が議会通过
→2018年に法律が施行



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が施行後も、個票データの利用に関する基本的な考え方は変わらない。
→5つの安全モデルによるデータ提供の枠組み

参考 イギリスにおける行政記録情報の二次利用

・Digital Economy Act 2017に基づいて行政記録情報を利用するにあたっては、**認定された研究者(accredited researcher)** になるための資格取得が求められる。

←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。

・研究プロジェクトが行政記録情報の利用申請を行うと、**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、**Research Accreditation Panel**による審査を受ける必要がある。

・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データの利用が可能になる。

⇒行政記録間のリンケージもONSが実施する←共通IDがないため、名前、住所、誕生日といった直接的な識別子を使用して、リンケージを行う。

・行政記録情報を用いて行われた分析の成果物に関する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、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。

⇒公的統計の個票データと同様に、**公益性を有する研究(public good research)**であることが行政記録情報の二次利用の要件となっている。

②ドイツにおけるマイクロデータの提供状況*

ドイツ連邦統計局におけるマイクロデータ提供の現状

1) Scientific Use File (SUF) の提供

→「事実上の匿名性(factual anonymity)」の概念に基づき、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

※事実上の匿名性・・・「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

⇒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

2) Public Use File (PUF) の公開

⇒「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) の考え方に基づくPUFやCampus Fileの提供

3) オンサイト施設による個票データの提供サービス

「形式的な匿名化された個票データ(formally anonymized individual data)」「(非識別データ)へのアクセス(連邦統計法第16条「秘密保護」第6項)

4) プログラム送付型のリモートエグゼキューション

⇒リモートアクセスは、データの移送にあたるために、法律上リモートアクセスが認められないことから、ドイツ連邦統計局では、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が展開

③アメリカにおけるマイクロデータの提供の現状*

1. 一般公開型マイクロデータサンプル(Public Use Microdata Sample=PUMS)および一般公開型ファイル(Public Use File=PUF)の提供

* HP上で無料でダウンロードすることが可能

Ex. 人口センサス(Population Census)や経常人口調査(Current Population Survey)

2. リサーチデータセンターによる個票データの利用サービス

・センサス局の経済研究センター(Center for Economic Study)を拠点として、全国33か所の大学等に連邦統計リサーチデータセンター(Federal Statistical Research Data Centers=FSRDCs、以下「リサーチデータセンター」)が設置され、FSRDCsにおいて、世帯・人口系だけでなく、事業所・企業系の公的統計の個票データだけでなく行政記録情報のアクセスが可能になっている。

←宣誓職員制度(Special Sworn Status)による個票データの利用

→ResearchDataGovによる個票データの検索システムの運用が開始されている。

⇒SAP(Standard Application Process) Portalにおいて、16の統計作成部局における個票データの検索が可能

3. リモートアクセスによる個票データの提供

・現在、リサーチデータセンターでの個票データの利用経験がある研究者を対象に、リモートアクセスによる個票データの利用が可能な仕組みが整備されている。

*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

④デンマークにおけるマイクロデータの提供状況*—デンマーク統計局の例—

・デンマークの場合、リモートアクセスを通じてのみ、個票データ(非識別データ、deidentified data)の利用サービスが行われている。

・デンマークにおけるマイクロデータ提供の経緯

1980年代半ば:デンマークにおける学術研究のための公的統計のマイクロデータの利用に対する関心の高まり

1988年:デンマーク統計局内にセキュアな環境であるオンサイト施設が設置

2001年:リモートアクセスサービスが開始

2008年:オンサイト施設は閉鎖され、リモートアクセスのサービスが拡充

*伊藤伸介(2020)「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

※利用可能なデータの一覧

①人口社会データ

結婚と離婚、出生と死亡、家族編成、移住、親子関係等

②労働、所得、社会データ

就業、失業、職業訓練、個人所得、賃金、家族所得、児童扶助、犯罪等

③教育データ

最終学歴、学年、教育歴、成人での教育等

④企業部門データ

企業の雇用、購入・販売、勘定統計等

⑤健康データ

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

- ・マイクロデータのアクセスに関しては、**認証された機関 (authorized institution)**においてのみ個票データの提供が可能

- ・認証された機関について

- ・大学等の機関が認証を受けるための要件

- ①安定した研究環境(stable environment)

- ②責任のある管理者がいること

- ③マイクロデータを扱った経験を持つ研究者がいること

大学等の機関が認証を受けるためにデンマーク統計局に申請をしなければならない

- ・**認証されている所属機関の研究者及びその共同研究者**であれば、認証機関から個票データにアクセスすることができる。

→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。

⇒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。

- ・個票データの利用を希望する研究者は、個票データの利用申請を行い、誓約書に署名した上で、利用することが可能

・「知る必要性(need to know)原則」に基づいて、利用者において必要なマイクロデータのみアクセスすることが可能

利用申請の時点で、デンマーク統計局の担当部局である研究サービス課 (Research Service Unit) が、分析に必要なデータを申請するように指導

→プロジェクトごとに別々のIDが振られており、そのIDにしたがって、デンマーク統計局がリンクージュを行う。そのため、研究者がIDに直接アクセスすることはできない。

・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)に基づく法的な根拠に基づいて、マイクロデータが提供される。

⇒2018年デンマークデータ保護法に基づいて、統計目的で集められた情報を他の統計目的や研究目的のために加工することが可能になっている。

※個人情報[※]を統計目的や学術研究目的で利用することはGDPRの適用除外という扱いになる。

分析結果のチェックについて

・分析結果のチェックについては、すべての分析結果ではなく、一部の結果をランダムに選んだ上で、職員が事後的にチェックを行う。

→閾値ルールは用いられている。

・審査担当者は、研究者が分析結果を持ち出す場合のデータの容量をチェックしており、容量が大きいデータについては、個票データあるいはそれに準じた高次元の集計表である可能性が高いと考えられる。

・基本的には、デンマークにおいては、分析結果のチェックに関しては統計作成部局と研究者との信頼関係が重視されている。

2018年に、デンマーク統計局は**分析結果のスキャンシステムを開発**

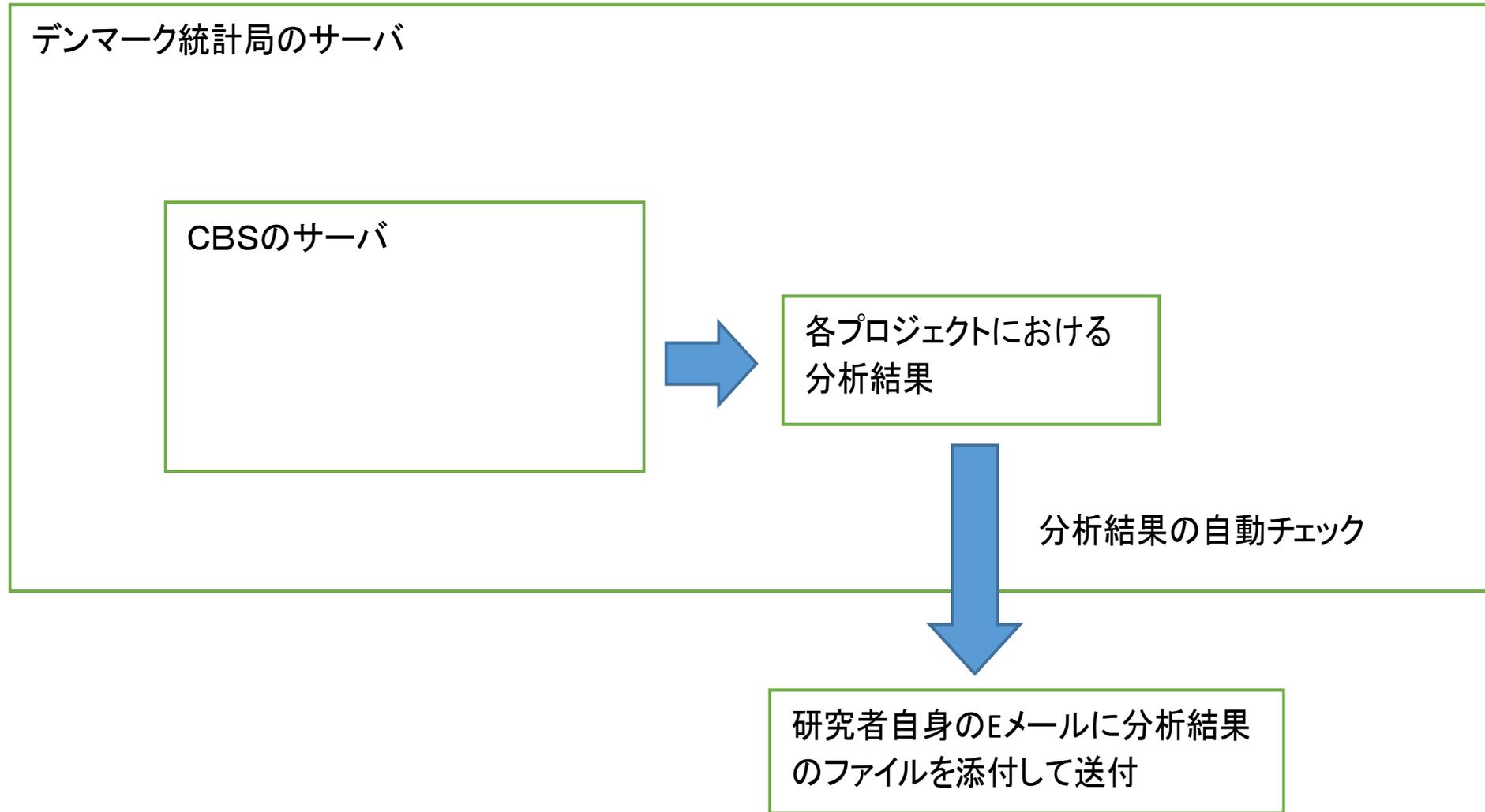
→研究者が分析結果のチェックを自動的に行うことが可能



・マイクロ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可能

→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した結果については、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成果物を載せることができる。

図4 デンマーク統計局と認証機関がアクセスできるサーバとの関係



注 コペンハーゲンビジネススクール(CBS)におけるヒアリング調査(2019年2月26日実施)をもとに作成

4. まとめ

・海外におけるマイクロデータの提供には、各国の法制度や利用者のニーズによって、多様な提供形態が存在するが、個票データについては、セキュアなオンサイト施設のもとで、公的統計や行政記録情報の個票データに関する利用サービスが行われているだけでなく、研究室からの個票データリモートアクセスも広範に展開されている。

⇒個票データの利用申請書類に関する審査が行われるが、利用承認後、個票データによる分析結果について、事後的な秘匿性のチェックを行うことも求められる。

・イギリスのように、研究プロジェクトの成果が公共の利益に資するかどうか、個票データの提供に関する判断基準となっている事例が存在する。

・個票データの利用の要件となる資格取得制度としては、例えば、研究者への利用認可制度(イギリス)、利用者が所属する研究機関に対する認証制度(デンマーク等)、宣誓職員制度(アメリカ)がある。

・個票データの一括検索を可能にするためのメタデータの整備と検索システムの開発がなされていることも興味深い。

⇒ONSのSRSで個票データを利用するためのカタログの整備やアメリカにおけるResearchDataGovによる個票データの検索システムの運用