

公的統計の二次的利用の現状について

2017年7月6日

中央大学経済学部

伊藤 伸介

1. はじめに

諸外国では、秘匿性と利用者のニーズを踏まえた形で、公的統計(政府統計)マイクロデータの提供に関して多様なチャンネルが存在する。

⇒公的統計データの提供状況は、個別具体的にみると、各国によって異なる様相を呈している。



本報告では、最初に、イギリスを中心に、諸外国における公的統計の二次的利用の現状とその方向性を明らかにする。つぎに、わが国における公的統計データの提供における将来的な方向性を探ることになりたい。

2. 欧米諸国における公的統計マイクロデータの作成・提供の現状

(1) 匿名化マイクロデータ(anonymized microdata)の特徴

①複数のファイルの提供

1)一般公開型ファイル(public use file) or 学術研究用ファイル(scientific use file)

Ex. ドイツ, イタリア, イギリス, オランダ etc.

* イギリスのライセンス型マイクロデータ(End User Licensed Data)は学術研究用ファイルとしての位置づけ。

* * オーストラリアやニュージーランドの匿名化マイクロデータファイル(Confidentialised Unit Record Files=CURFs) は学術目的用ファイルとしての位置付け(Basic CURFs, Expanded CURFs)

* * * 一般公開型ファイルについては,特定の目的に対する利用を想定したマイクロデータが作成されているだけでなく(ex. イギリスの教育用マイクロデータ),国によっては提供形態が異なる場合がある。

・webからのダウンロードによる提供

アメリカ,イギリス etc.

・有料での一般公開型ファイルの提供

カナダ etc.

2)世帯単位で抽出or 個人単位で抽出←世帯・人口系のデータ

・世帯単位で抽出されたマイクロデータファイルは、階層型ファイルになっており、世帯構成員の属性が利用可能であるのに対して、個人単位で抽出されたマイクロデータファイルは、地域区分が詳細

Ex. アメリカ, イギリス, カナダ etc.

* オーストラリアの場合, 世帯単位で抽出された抽出率の異なる(1%, 5%)人口センサスのマイクロデータ(Census Sample Files (CSFs))が提供

3)データの種類

・世帯・人口系のデータ or 事業所・企業系のデータ

⇒一般に, 人口センサスや労働力調査等, 世帯・人口系の調査について匿名化マイクロデータが作成されている。

* イタリアの場合, 企業データであるItalian innovation surveyが一般公開型ファイルと学術研究用ファイルの両方で提供されている。

* * Eurostatの場合, 企業データであるCommunity Innovation Surveyが, 学術研究用ファイルとして提供されている。

4)データの提供形態

・CD-ROM or ダウンロード←学術研究用ファイル

CD-ROM: オーストラリア, ニュージーランド, カナダ, オランダ, Eurostat etc.

ダウンロード: イギリス, イタリア etc.

②匿名化マイクロデータの作成方法(Zayatz(2007), 伊藤(2016c))

匿名化マイクロデータの作成においては,データの特性や秘匿の程度に応じて,様々な匿名化技法が適用される。

非攪乱的な手法

- ・データの削除(レコード削除,変数の削除,suppression)
- ・リコーディング(recoding)
- ・トップ(ボトム)コーディング(top(bottom)coding)
- ・地理的な閾値の設定
- ・分類区分の閾値の設定

攪乱的手法(パータベーション)

- ・スワッピング
- ・ノイズの付加
- ・丸め(ラウンディング)
- ・PRAM(=Post RAndomization Method)

* オランダの場合, μ -Argusを用いて, public use microdata files が作成されている(Nordholt(2013))

* * Eurostatでは,合成データ(synthetic data)の方法論を用いてpublic use dataの作成に関する研究が行われている(de Wolf(2015))

(2) オンデマンド型の提供システム(remote execution)

1) オンデマンド型の提供による出力結果

- ・リモート集計のみ

 - StatLine(オランダ), TableBuilder(オーストラリア)等

- ・モデル分析も可能

 - NESSTAR(イギリス等)

2) 保管されるデータのタイプ

- ・集計表 or data cube(ex. StatLine(オランダ), American FactFinder(アメリカ))

- ・マイクロデータ(ex. NESSTAR(イギリス等), Data Ferrett(アメリカ), TableBuilder(オーストラリア)等)

(3) オンサイト施設による個票データのアクセス (on-site access)

1) オンサイト施設の場所

- ・統計作成部局(イギリス,カナダ,オーストラリア,ドイツ等)
- ・大学(リサーチデータセンター)(アメリカ,カナダ等)
- ・その他研究機関(ex, IAB(ドイツ))

2) 利用可能な個票データの数と種類

- ・世帯・人口系のデータだけでなく事業所・企業系のデータもアクセス可能
⇒事業所・企業系のデータについては,統計作成部局内のオンサイト施設でのみアクセス可能な国も存在(例えば,カナダ等)
- ・行政記録データのアクセス可能性(ex. イギリスのADC(=Administrative Data Centre))
- ・統計調査間さらには統計調査と行政記録データにおけるリンケージの可能性

(4) リモートアクセス(remote access)

1) リモートアクセスに基づく個票データの利用

- ・プログラム送付集計型

ex. ルクセンブルク所得調査(Luxemburg Income Study)

- ・直接利用型

ex. イギリス, ドイツ, フランス, イタリア, オランダ, スウェーデン,
デンマーク, フィンランド, スロベニア, オーストラリア,
ニュージーランド等

2) どの組織が管理するのか

- ・統計作成部局(ex. オランダ, オーストラリア等)

- ・大学(ex. イギリス)

3) 利用可能なデータの種類

- ・世帯・人口系のデータ

- ・事業所・企業系のデータやビジネスレジスターが利用可能な
国も存在(ex. イギリス)

3. イギリスにおける公的統計の二次的利用について (伊藤(2011, 2012, 2014, 2016b))

(1) ライセンサ型マイクロデータ(End User Licence Data)の提供

⇒イギリスの代表的なデータアーカイブ機関であるThe U. K. Data Archive(UKDA)(エセックス大学)のUK Data Serviceにおいて提供サービスが行われている。

→労働力調査(LFS)等のサーベイマイクロデータ, 人口センサスのマイクロデータ等

→オンデマンド型システムであるNESSTARを通じてアクセスすることも可能

(2) Public Use Fileの公開

⇒2011年人口センサスについては, 人口センサスの教育用マイクロデータが提供

(3) オンサイト施設による個票データの提供サービス

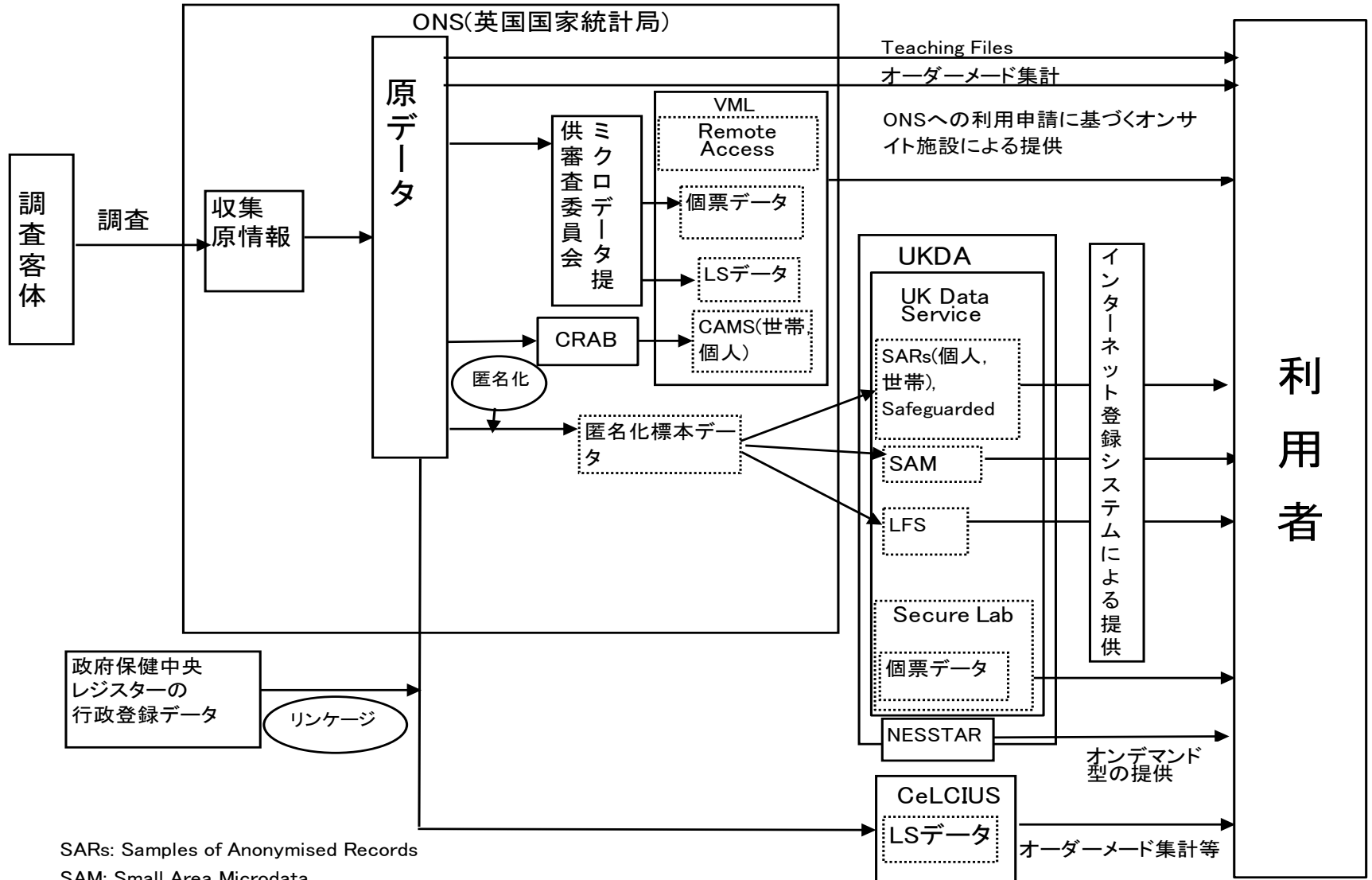
⇒イギリス国家統計局(The Office for National Statistics=ONS)のVirtual Microdata Laboratory(VML)のようなオンサイト施設において, 公的統計の個票データが利用可能

* LSデータ(ONS Longitudinal Study of England and Wales)もVMLにおいてアクセスすることができる。

(4) リモートアクセスに基づく個票データのアクセス

⇒ UK Data ServiceのSecure Labを通じて, リモートアクセスを行うことができる。

図 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



SARs: Samples of Anonymised Records

SAM: Small Area Microdata

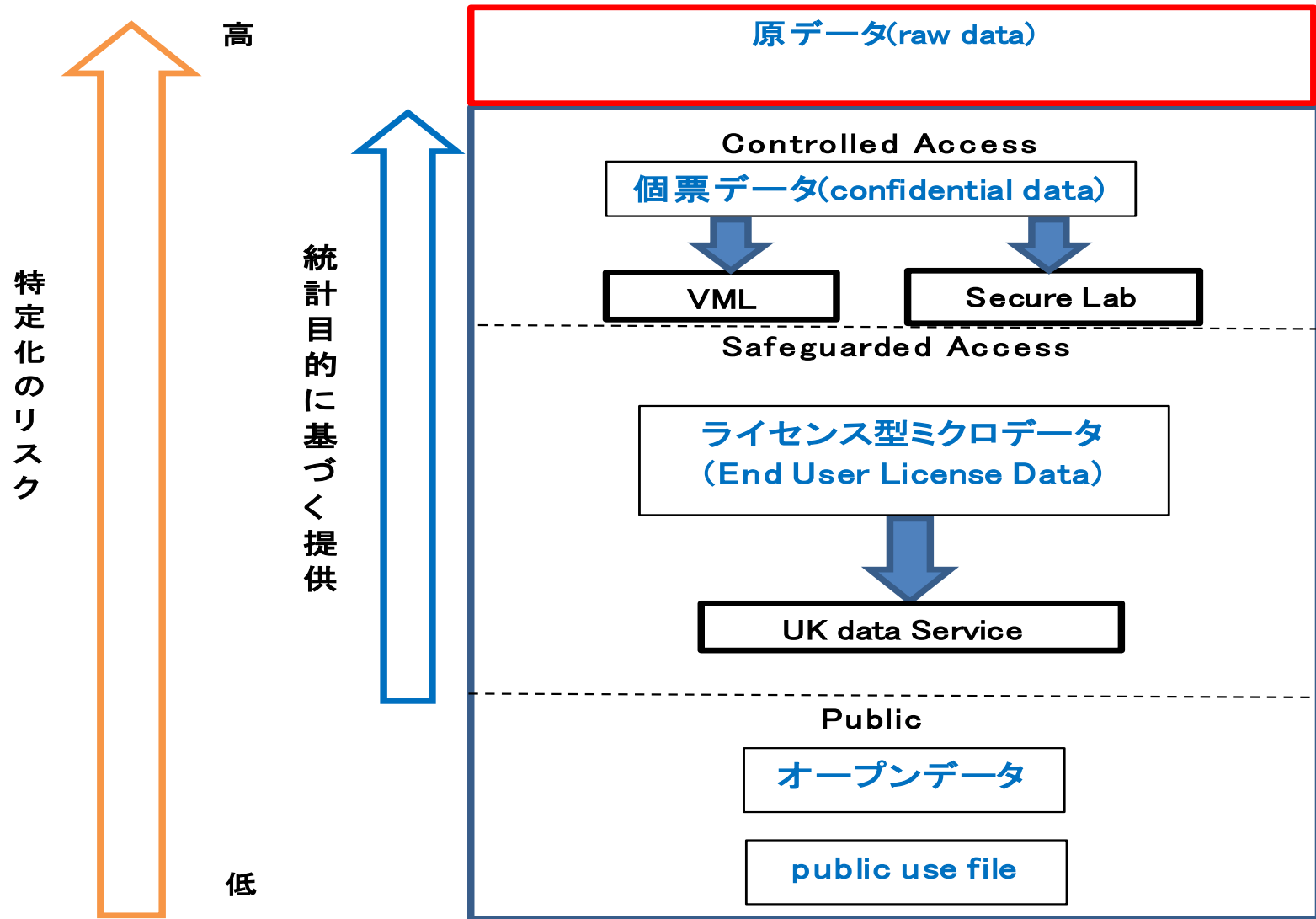
LFS: Labour Force Survey

CeLCIUS: Centre for Longitudinal Study Information and User Support

CAMS: Controlled Access Microdata Sample

CRAB: Census Research Access Board

参考 イギリスにおけるマイクロデータの提供の位置付け



参考 イギリス統計法第39条 個人情報秘密保護

(伊藤(2012))

(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、委員会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)

「承認された研究者」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。

・**研究目的**

・個票データを利用する期間

・**個票データの利用経験**

・申請者の所属機関

・研究成果について公表する論文

* 個票データの利用申請は、原則イギリス国内の研究者に限定されている。大学院生は、個票データを使用することはできるが、プロジェクトに指導教員が関わっている必要がある。

⇒利用を希望する研究者は、特定の学術研究のプロジェクトに所属していることが求められる。

* イギリスでは、個票データの申請において具体的な集計様式や分析モデルを記載することは求められない。

←LSデータ以外の個票データについては、分析のために最低限必要な属性ではなく、調査票に含まれるすべての属性を利用することが可能。

オンサイト施設による個票データの提供

—VML—

- ・ Virtual Microdata Laboratory(VML)は2004年にONSに開設 (Ritchie(2008))

⇒現在、ONSのロンドン本部、ティッチフィールド支局とニューポート支局を含む5か所に設置

- ・VMLにおいて、利用者は原データ(raw data)にアクセスすることは不可能であり、個票データ(disclosive data, confidential data)にしかアクセスすることができない。

- ・VMLのなかに、オンサイト施設であるsafe roomとリモートアクセス施設が存在

⇒イングランド銀行を含む政府機関しか、VMLのリモートアクセス施設を使用することができない。

- ・民間の研究者については、統計目的であればVMLで利用することができる。政府機関の研究者も同様。

- ・ONSが実施している統計調査の個票データがアクセス可能

リモートアクセスによる個票データのアクセス

—Secure Lab—

エセックス大学のU.K.データアーカイブにおいて、ESRC(=Economic and Social Research Council)の資金提供を受けたthe Secure Data Service(=SDS)というリモートアクセス施設が創設⇒2011年にその運用が開始

・SDSは、UKDAにおいてデータの寄託、利用者の登録・管理、さらにはデータの提供サービスを担ってきたESDS(=Economic and Social Data Service)等と統合

→2012年にUK Data Serviceが創設

⇒現在は、UK Data Service のthe Secure Labがリモートアクセスのサービスを行っている。

Five Safes modelについて

(Ritchie(2008),Desai *et al.*(2016))

Five Safes Model:2003年に考案されたモデルで,ヨーロッパだけでなく,オーストラリアでも現在適用されているモデル

*** 個票データ(confidential data)のアクセスは,以下の5つの基準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見えるような結果が含まれないこと。



安全な利用(safe use)

安全な利用者(safe people)

個票データのアクセスにおいて,利用者に対する「信頼」を重要視している。

・イギリスの場合,ONSが実施している統計調査の個票データを利用するためには,「承認された研究者(approved researcher)」の資格を取得する必要がある。

・個票データにアクセスするためには,個票データの利用に関する法的根拠,利用手続き,分析結果の秘匿処理について特別のトレーニングコースを受けることが求められる。

⇒ 個人情報を漏えいした場合の罰則規定に関する説明の
必要性

安全な分析結果(safe outputs)

・分析結果に関する最終成果物(final output)については,ONSの Virtual Microdata Laboratoryの担当職員やエセックス大学の Secure Labの担当者が秘匿性に関するチェックを行った上で公表することが可能になる。

⇒Principles-Based Approachの適用

- Rule of Thumbモデル

- 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

- 分析結果のチェックに関する自動化は行わない。

▪ 個票データの提供者側と利用者側の協力のもとで秘匿性が確保されるだけでなく,それに基づきながら,分析結果が公共の利益(public benefit)に資することが可能な形で個票データのアクセスが行われる。

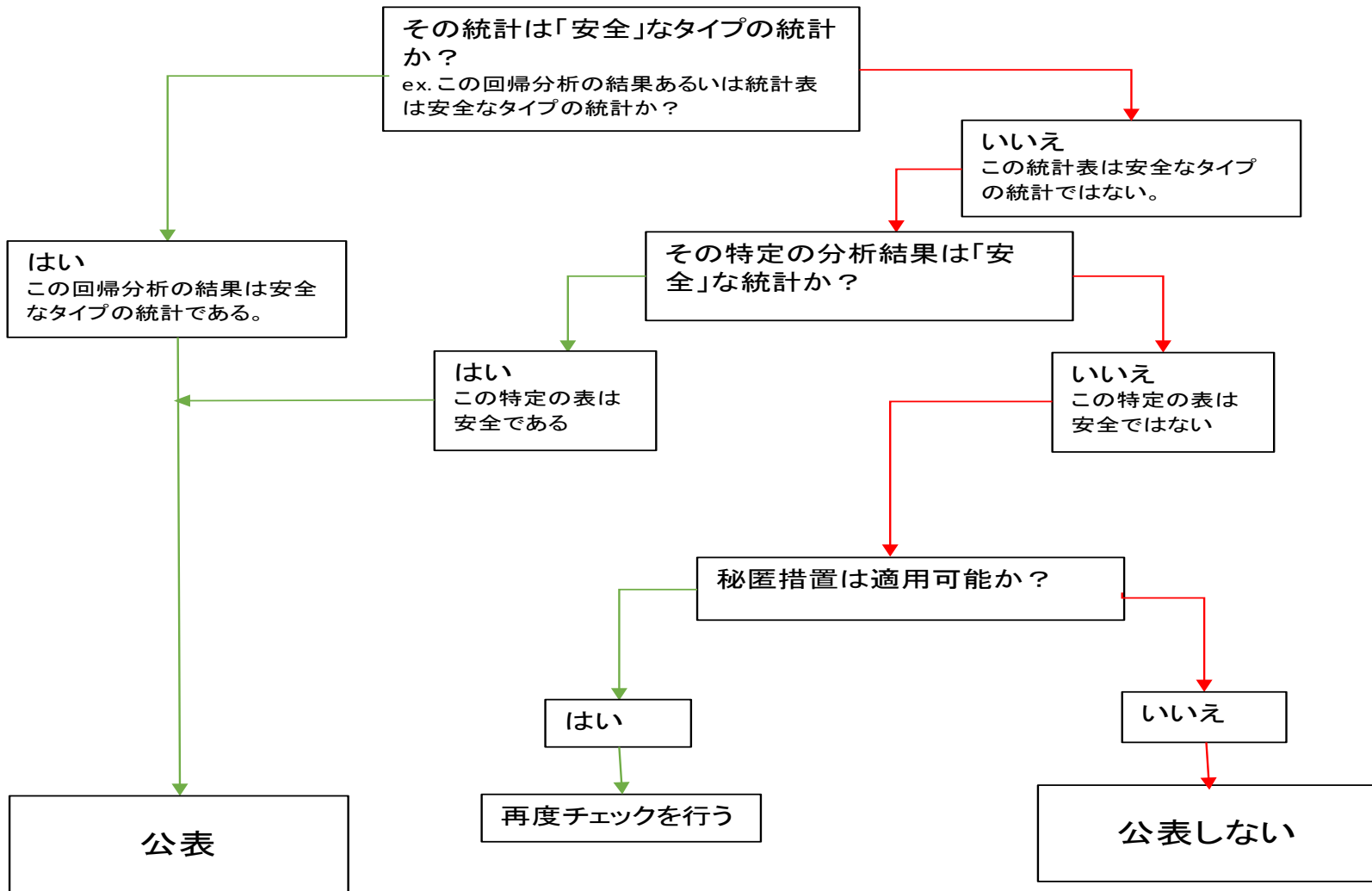
「安全な」統計について

分析結果のタイプによって「安全な」タイプの統計が「安全でない」タイプの統計を類別することが可能

| | | |
|-----------|---|-----------|
| 統計のタイプ | 分析結果のタイプ | 安全か安全でないか |
| | 度数表(Frequency tables) | 安全でない |
| | 数量表(Magnitude tables) | 安全でない |
| | 最大値, 最小値, パーセンタイル(メディアンを含む) | 安全でない |
| | 最頻値 | 安全 |
| | 平均値, 指数, 比率, 指標 | 安全でない |
| | 集中度(Concentration ratios) | 安全 |
| | 分布に関する高次のモーメント(分散, 共分散, 尖度, 歪度を含む) | 安全 |
| | グラフ, 実際のデータに関する絵入りの表現 (pictorial representation) | 安全でない |
| 相関係数と回帰分析 | 線形回帰モデルにおける回帰係数 | 安全 |
| | 非線形回帰モデルにおける回帰係数 | 安全 |
| | 推定値の残差 | 安全でない |
| | 推定値に関する要約統計量および検定統計量(R^2 , カイ2乗値等) | 安全 |
| | 相関係数 | 安全 |

出所 Brant et al.(2010)

図 「安全な」統計の決定に関するチャート(decision tree)



出所 Ritchie and Welpton(2015)

4. わが国における公的統計データの提供と匿名化措置

公的統計データの提供形態

⇒データの秘匿性と利用者のニーズに合わせた形で様々なチャンネルが存在

- ①統計表(集計結果表)の公表およびオープンデータ化
- ②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)の提供
- ③個票データの提供
- ④オーダーメイド集計

わが国においては、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、公的統計のマイクロデータの提供が進められてきた。



わが国における公的統計データの提供状況

①統計表(集計結果表)およびオープンデータ

⇒e-Statによる統計表の公表, API機能の整備や統計GISの整備

②匿名化マイクロデータ

⇒「匿名データ」(統計法第2条第12項)として提供

→統計法第36条

・国勢調査等, 世帯・人口系の7調査に関する匿名データが現在提供されているが, 1種類の匿名データしか提供されていない。

③個票データ

⇒「調査票情報」(統計法第2条第11項)として提供

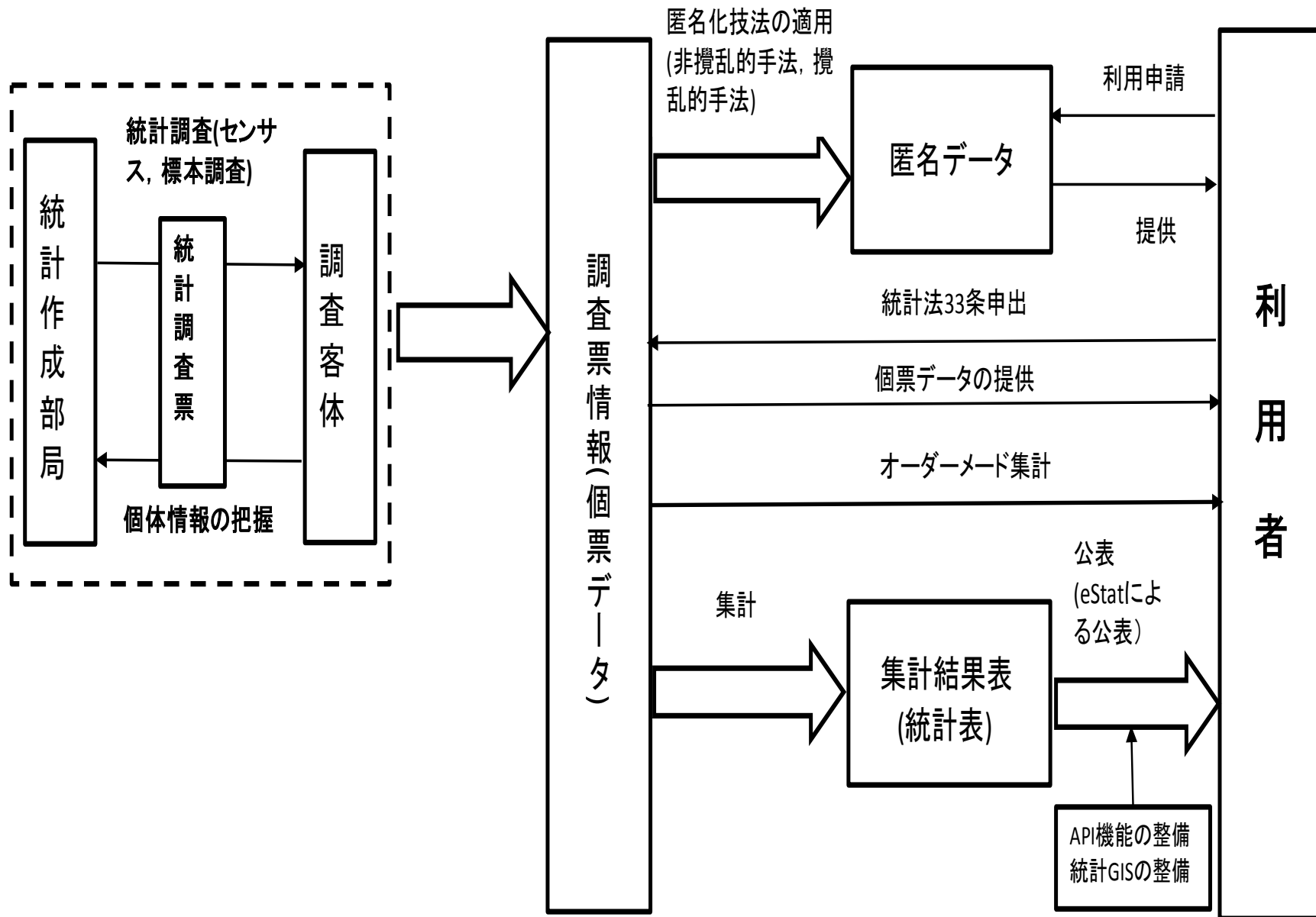
→統計法第33条

④オーダーメイド集計

⇒様々な統計調査を対象に有料でオーダーメイド集計が可能

→統計法第34条(委託による統計の作成等)

図 わが国における公的統計データの提供形態に関する概略図



出所 伊藤(2016a)

参考 統計法におけるマイクロデータの提供に関する条文 (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。

- (1)行政機関は、**統計委員会**へ匿名データ作成について諮問を行う。
- (2)統計委員会内に設置されている**匿名データ部会**において、基幹統計調査の匿名データに関する匿名性の検討が行われる。匿名データ部会の審議内容は、統計委員会に報告される。
- (3) 統計委員会において基幹統計調査の匿名データの作成に関する答申が出されると、行政機関は匿名データの作成・提供を行う。

(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。

⇒**調査票情報や匿名データの適正管理(第四十二条)、守秘義務(第四十三条)、罰則規定(第六十一条)あり。**

匿名データの作成・提供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)(以下『ガイドライン』)

(1)匿名データ提供の申出者および利用者の範囲(12頁)

- ・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

⇒学術研究や高等教育の発展に資すると認められる場合に限り、匿名データの提供が認められる。

←利用目的の明示

(2)匿名データの申出者は、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必要がある。

←利用する場所の指定

*わが国の匿名データは、Public Useな匿名化マイクロデータではない←オープンデータではない。

匿名データにおける匿名化処理の考え方(『ガイドライン』別紙1)

匿名化処理:「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」こと

⇒各種匿名化技法(『ガイドライン』別紙2)を組み合わせる匿名化処理を行う。

『ガイドライン』で記載されている匿名化技法

- ・ 識別情報の削除
- ・ 匿名データの再ソート(配列順の並べ替え)
- ・ 識別情報のトップ(ボトム)・コーディング
- ・ 識別情報のグルーピング(リコーディング)
- ・ リサンプリング
- ・ ミクロデータのソート
- ・ **スワッピング**
- ・ 誤差の導入

* 国勢調査の匿名データの作成において,初めてスワッピングが用いられる。

* * 主として外観識別性の高い属性に対して,匿名化処理が適用されると考えられる。

統計法第33条の運用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)

- ・調査票情報の作成・提供に関する手続き上のルールが詳細に記載されている。
- ・現行のガイドラインのもとでは、政府統計の個票データの利用申請を行う場合、実証分析を行う上で最低限必要な調査事項(変数)のみの利用が想定されている。
- ⇒ 個票データの提供については、オンサイト施設による利用に関する検討が進められている。
 - 事後チェック型の調査票情報の利用(小林(2012))に基づくさらなる探索的な研究の可能性
- ⇒ 分析結果の秘匿性のチェックに関する具体的な基準の必要性

5.おわりに

- ・本報告では、諸外国における公的統計マイクロデータの提供状況を明らかにした上で、わが国の公的統計における二次的利用の現状を明らかにした。

- ・諸外国では、匿名化マイクロデータの提供、オンサイト施設やリモートアクセスによる個票データの利用サービス等、複数のチャンネルによる公的統計のマイクロデータの提供が進められている。

⇒わが国の公的統計のマイクロデータに関しても、秘匿性のレベルに留意しつつ、利用者のニーズも踏まえた上で、データ提供に関する様々な形態が模索されるものと思われる。

- ・諸外国においては、学術研究を指向した特定の利用者限定したデータの提供(ex.オンサイト施設やリモートアクセスによる利用等)とオープンな形で利用可能なデータ(オープンデータ, public use file等)の公開は、明確に区別した上で行われている。

⇒わが国の匿名データにおける今後の位置付け

- ・海外の統計作成部局の中には、個票データの分析結果のチェックにおいて、有用性と秘匿性の両面からprinciple-based approachの適用可能性を追究している国が少なくない→わが国のオンサイト利用のあり方を議論する上でも参考になりうる。

⇒分析結果のチェックを行う担当者の審査技術の養成だけでなく、個票データを分析する研究者の側においても秘匿措置に関する基本的な知識の習得が必要になるだろう。

主要参考文献

- 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頁
- Brandt, M., Franconi, L., Guerke, C., Hundepool, A., Lucarelli, M., Mol, J., Ritchie, F., Seri, G., Welpton, R. (2010) *Guidelines for the Checking of Output Based on Microdata Research*, Final Report of ESSnet Sub-Group on Output SDC, Eurostat
- Desai, T., Ritchie, F., Welpton, R.(2016) “Five Safes: Designing Data Access for Research”, Economics Working Paper Series 1601, University of the West of England.
- De Wolf, P. P. (2015) “Public Use Files of EU-SILC and EU-LFS data”, Paper presented at Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Helsinki, Finland, pp.1-10.
- 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』第23巻第3号, 36～54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)「政府統計におけるリモートアクセス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, 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
- 小林良行(2012)「公的統計マイクロデータ提供の現状と展望: 一橋大学での取り組みをもとに」『日本統計学会誌』第41巻, 第2号, 401～420頁

主要参考文献

- Lauger A., Wisniewski, B., McKenna, L. (2014) “Disclosure Avoidance Techniques at the U.S. Census Bureau: Current Practices and Research”, Research Report Series (Disclosure Avoidance #2014-02), U.S. Census Bureau, pp.1-13.
- Nordholt, E.S.(2013) “Access to Microdata in the Netherlands: from a Cold War to Cooperation Projects”, Paper presented at Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Ottawa, Canada, pp.1-11.
- 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.
- Ritchie, F. and Weltpn, R. (2015) “Operationalizing principles-based output statistical disclosure control”, mimeo
- 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.