

第7回統計委員会と統計利用者との意見交換会

イギリスにおける政府統計データの 二次的利用の現状

2013年3月28日

明海大学経済学部

伊藤 伸介

1. はじめに

イギリスでは、人口センサスの匿名標本データ(以下「センサスマイクロデータ」と呼称)、標本調査のマイクロデータ(以下「サーベイマイクロデータ」と呼称)等、様々なタイプの政府統計マイクロデータが提供されている。

本報告の目的

イギリスにおける政府統計データの二次的利用の現状を把握することは、わが国における統計データの二次的利用に関する将来展望を図る上で、有益であることから、本報告では、人口センサスのミクロデータを中心に、イギリスにおけるミクロデータの提供状況を明らかにする。

イギリスにおいて利用可能な政府統計ミクロデータ

1) センサスミクロデータ (SARs, SAM等)

⇒センサス調査研究センター(CCSR)

2) サーベイミクロデータ (ex. 労働力調査(LFS)等)

⇒The U.K. Data Archive(UKDA)

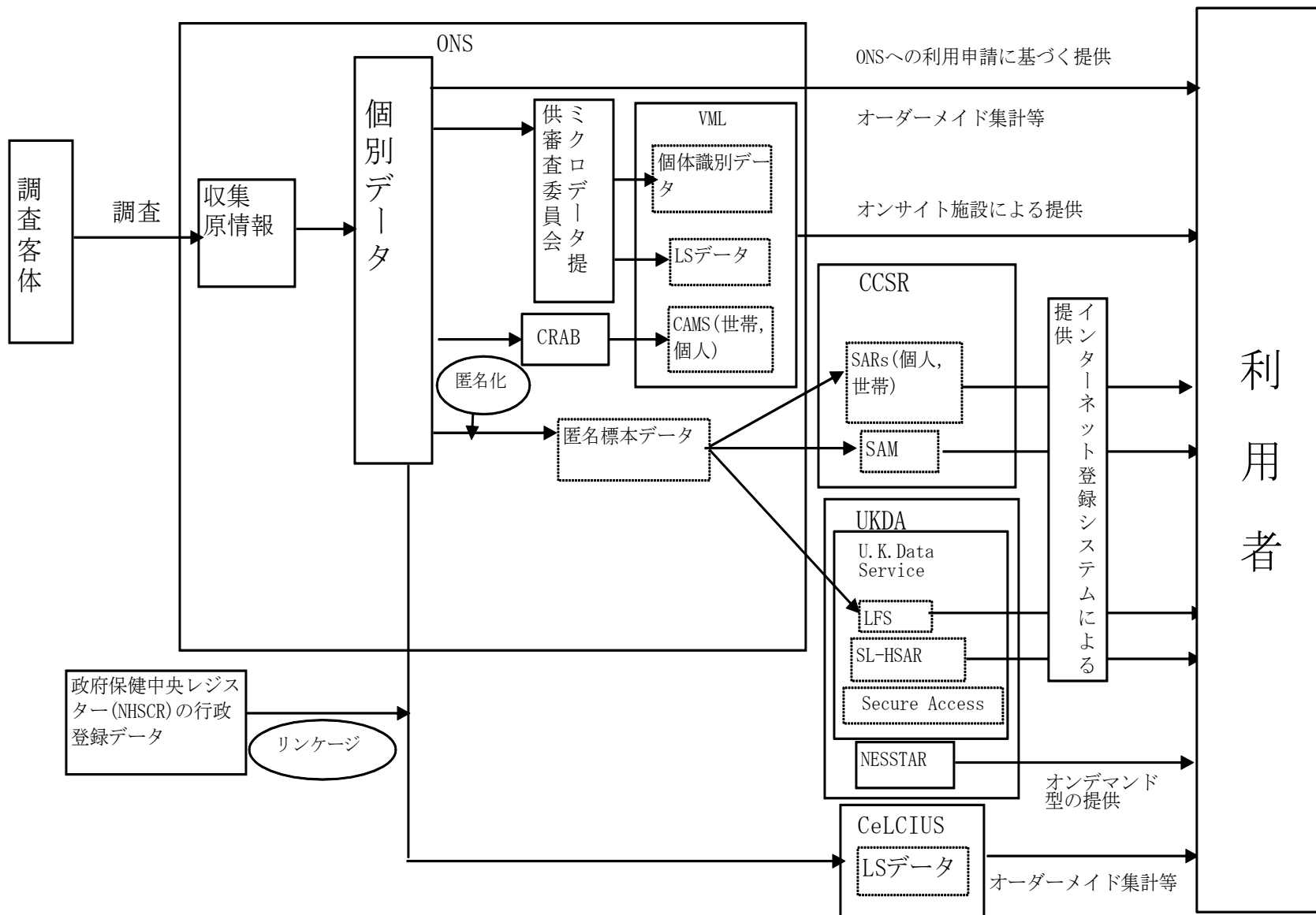
3) LSデータ (ONS Longitudinal Study of England and Wales)

⇒Centre for Longitudinal Study Information and User Support(CeLSIUS)

4) 個体識別データ (identified data)

⇒イギリス国家統計局(ONS)

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



注 伊藤(2011, 54頁)をもとに作成

2. センサスマイクロデータの提供

(1)匿名化標本データ (Samples of Anonymised Records=SARs)

(2)小地域マイクロデータ (Small Area Microdata=SAM)

(3)Controlled Access Microdata Sample(CAMS)

2-1 匿名化標本データ (SARs)

人口センサス：個人の人口学的属性(性別，結婚状態等)や就業属性(産業，職業等)，さらには家族属性(家族類型等)，世帯属性(住居の類型等)に関する様々な属性が捕捉される。

⇒細かな地域区分を有する**個人SAR**と個人・世帯属性を有する**世帯SAR**が提供されている。

→人口学，社会学，地理学といった様々な分野におけるセンサスミクロデータに対する要望に応えるため。

SARsの特徴(森(2000), 伊藤(2005), 伊藤(2011))

(1)1991年SARs)

1)世帯SAR

- 標本抽出率は1%(約50万の標本数(約216000世帯))
- 詳細な個人・世帯属性
- 世帯構成員間のリンケージが可能なように、世帯の階層性(hierarchy)が保持
- 地域区分については、グレートブリテン島全域において12地域に限定

2)個人SAR

- 標本抽出率は2%(約110万の標本数)
- 個人・属性の分類区分は粗い
- 地域区分に関しては、最低12万人の人口を有する278の詳細な区分が利用可能(伊藤(2005, 37頁))
- 年齢は90歳までは各歳年齢区分で提供

⇒地域分析に利用可能

＊1991年SARsには導出変数の形でウェイトが作成されている←CCSRが設定

(2)2001年SARs

1) Special Licence型世帯SAR(Special Licence Household SAR=SL-HSAR)

- 標本抽出率は1%(約50万の標本数(約20万世帯))
- 対象地域はEngland and Walesのみ
- イギリス国家統計局からのライセンスが必要

2)個人SAR(Individual SAR(Licensed))

- 標本抽出率は3%(約184万の標本数)
- 地域区分はgovernment office regionのレベルでのみ利用可能

***2001年SARsにはウェイトは存在しない。**

SARsにおけるレコードの抽出方法(森(2000))

SARはセンサスに含まれる全レコードの中の完全コード化レコード(fully coded records, 全レコードの10%)から作成される。

⇒世帯SARの対象となる世帯レコードを先に抽出し、その残りのレコードから個人SARに含まれる個人のレコードが抽出

(1)1%世帯SAR(England & Wales)

最初に、州、調査区ごとに層化し、10世帯の組を編成した上で、世帯の組ごとに1世帯を抽出する。つぎに、地域が特定されないように地域間のレコードの入れ替えを行う(scramble)。

(2)2%個人SAR(グレートブリテン島地域)

最初に、完全コード化レコードから世帯SARの抽出対象レコードを除いたレコードにおいて、住居に居住する個人については9人ずつ、公共施設の居住者については5人ずつの組を編成する。これらの組の中から、前者については2人ずつ、後者については1人ずつランダムに抽出し、地域間でレコードを混ぜ合わせる。

SARsに適用された匿名化技法

1991年SARsにおける匿名化技法

1) レコードの標本抽出

2) データの削除(suppression)

3) 地域属性の制限

⇒地域属性については、世帯SARが12地域，個人SARが120,000レコード以上の地域に区分

4) 個人・世帯属性の分類区分の統合

⇒世帯SARについては2,700レコード，個人SARに関しては25,000レコードの閾値がそれぞれ設定

⇒イギリスの人口センサスミクロデータでは，非攪乱的手法が推奨されている。

・ 2001年SARsの作成においては、攪乱的手法 (perturbation)として、PRAM(Post Randomization Method)が適用されている(De Kort and Wathan(2009))。PRAM: ミクロデータにおける属性値に対して、事前に設定されたマルコフ連鎖遷移行列に基づいて攪乱を行うこと(藤野・垂水(2003))
⇒「特殊な一意(Special Uniques)」を処理するため

・ イギリスでは、人口センサスの個別データの作成において、レコードスワッピングが適用されている(Shlomo(2007))。

SARsの入手方法

(I)1991年のSARsおよび2001年個人SAR

- ・ イギリス国内において学術目的で利用する研究者・学生がU.K. Data Serviceにおいて利用者登録を行い、End User License Agreementに同意すれば、学術目的で利用する教育機関の研究者・学生は、Web上でセンサスマイクロデータを無料でダウンロードすることができる(以下「ライセンス型マイクロデータ (End User License data)」と呼称)。

⇒ Safe data

- ・ NESSTARのようなオンデマンド型の提供システムを通じて、センサスマイクロデータを利用することも可能←データキューブではなく、ライセンス型マイクロデータが保管されている。

←NESSTARについてはモデル分析を行うことも可能

(2)Special Licence(SL)型の世帯SAR

ONSからのライセンス(Special Licence)を取得することによってアクセスが可能←「承認された研究者
(approved researcher)」のみが利用可能

2-2 小地域マイクロデータ (SAM)

小地域マイクロデータ (SAM)

- 標本抽出率は5%(約300万の標本数)
- ← 個人単位で抽出されており，SARsと重複しないようにレコードの系統的な抽出が行われている。
- 地方自治体(Local Authority)レベルの細かな地域区分が利用可能
- 個人・世帯属性については，SARsと比較して粗い分類区分を持つ属性やレコードから削除される属性が存在
- SAMもライセンス型マイクロデータとして取得可能

表 SARsおよびSAMに含まれる主な属性

	個人SAR (Licensed)	Special Licence型世帯SAR	小地域マイクロデータ (SAM)
サンプル数	1843525	525715	2964871
属性	分類区分の数		
年齢	15歳までは各歳年齢階級 16～19歳 20～24歳 25～29歳 30～44歳 45～59歳 60～64歳 65～69歳 70～74歳 75～94歳までは各歳年齢階級 95歳以上は区分統合	0～1歳 2～3歳 4～5歳 6～7歳 . . . 76～77歳 78～79歳 80歳以上	0～4歳 5～9歳 10～15歳 16～19歳 20～24歳 25～29歳 30～39歳 40～49歳 50～59歳 60～64歳 65～74歳 75～84歳 85歳以上
出生国	16	16	5
民族グループ(イングランドとウェールズ)	16	16	13
1週間の就業時間	80	80	5
産業 (SIC)	18	17	削除
職業 (標準職業分類)	81	81	削除
通勤手段	12	11	7
家族類型	9	9	4
住居の類型	7	7	3

導出変数について

- ・ 個人SARや世帯SARにおいては導出変数が含まれるが、SAMには導出変数が設定されていない。
- ・ 導出変数の中にはONSで作成される変数と、CCSRで新たに設定される変数が存在する。

2-3 CAMS

CAMS(Controlled Access Microdata Sample):

ONS内部のオンサイト施設でのみ利用可能なセンサス
ミクロデータ←「**個人情報(personal information)**」としての位置付け

⇒2001年センサスについて個人CAMSと世帯CAMSの2
種類が提供←**SARsと同じ抽出率でCAMSが作成**

例えば、個人CAMSについては、人種・民族(16区分)、年齢（1歳ごとに95歳まで区分）、職業(SOC(Standard Occupation Classification) minor plus SOC Unit Codes)、産業（60区分）等の調査項目に関する詳細な分類項目が得られる。

⇒CAMSを取得するためには、ONS内部のCensus Research Access Board(CRAB)等による審査を受けることが要求される。

3. サーベイマイクロデータの提供

1970年代よりイギリス最大のデータアーカイブ施設である**UKDA**(当時はthe Survey Archive)によってサーベイデータが提供

⇒約6,000(2007年時点)にわたるサーベイデータが保管

提供されている主な政府統計のサーベイマイクロデータ

- ・ 労働力調査(Labour Force Survey)
- ・ 一般生活調査(General Lifestyle Survey 以前のGeneral Household Survey)

など

* イギリス国内において学術目的で利用する研究者・学生は,
U.K. Data Serviceを通じて無料でダウンロード可能

→NSSSTARによるオンデマンド型の利用も可能

4. LS データの提供

- ①(1)職業と死亡の関連性に対する社会的関心, (2)慢性病による死亡とその原因の追究, (3)女性の出生行動に関する傾向的な把握といった実証分析に必要なデータが得られなかったことから、1960年代末にLSデータの作成計画が進められ、1974年に人口センサス調査局(OPCS)がLSデータの作成に着手(森(2003, 28～29頁))
 - ②人口センサスと政府保健中央レジスター(NHSCR)の行政登録データをリンケージすることによって作成
 - ③イングランドとウェールズにおける約50万のセンサス個票レコード(individual-level Census records)を1971年～2001年の4時点についてリンケージ
- 1971年のセンサスレコードを更新するだけでなく、調査時点における出生者、移住者等が標本レコードに追加され、死亡者等がレコードから削除

・ LSデータにおいては、個人が特定される可能性が非常に高いことから、その提供においては、多くの申請手続き、ミクロデータ提供審査委員会(Microdata Release Panel)等の審査が要求されるだけでなく、利用に対しても厳しい制約条件が課せられている
⇒オンサイト施設での利用やオーダーメイド集計等

⇒イギリス国内の学術機関における教職員・学生に関しては、LSデータが無料で利用可能

5. イギリス国家統計局(ONS)による個 体識別データの提供

個体識別データは、イギリス国家統計局内部
のオンサイト施設(Virtual Microdata Laboratory)
で利用可能

⇒個体識別データに対する高いニーズ

←「承認された研究者(approved researcher)」のみが利
用できる。

←イギリスの統計法で規定

イギリスの統計法における秘密保護(伊藤(2012))

2007年に統計登録サービス法(The Statistics and Registration Service Act 2007(SRSA), 以下「統計法」と呼称)が公布→2008年4月に施行

- 「個人情報(personal information)」の秘密保護に関する条項の明記
- 統計法における「承認された研究者」の位置付け

統計法第39条 個人情報秘密保護

(1)この条項にしたがって、統計委員会によって保有される個人情報、委員会における職務の行使を通じて、以下の者によって開示されてはならない。

(a)委員会の委員ないしは職員

(b)委員会内の部会委員

(c)委員会から直接ないしは間接情報を受け取った委員

(2)この条文における「個人情報」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)

個人情報 の例

- 個体識別データ
→ ONS内部のVirtual Microdata Laboratory(VML)において利用可能
⇒ Safe settingsかつSafe people
- Special Licence型のデータセット(ONS)
→ 2001年人口センサスの世帯SAR, 労働力調査等に適用

*** ライセンス型マイクロデータは、個人情報ではない。**

出所 *GSS Disclosure Control Policy for Microdata Produced from Social Surveys*

個人情報を利用可能になるためには、「承認された研究者」の資格を取得する必要がある。



申請書類の提出



- National Statisticianによる承認
- ミクロデータ提供審査委員会による承認

6. おわりに

本報告では、イギリスのセンサスミクロデータを中心に、ミクロデータの提供状況の特徴を追究した。

- センサスミクロデータについては、標本抽出率、地域区分および個人・世帯属性の分類区分が異なる2種類のSARs, SAMとCAMSが作成・提供されている。
- 2007年統計法の施行に伴い、「個人情報」の秘密保護が条文に明記されることによって、「承認された研究者」による政府統計の個体識別データの利用可能性が法的に整備された。→個体識別データのさらなる提供

←ライセンス型ミクロデータは、「個人情報」の対象範囲の外に位置付けられることから、ライセンスの取得による法制度的措置と匿名化技法の適用による個体情報の匿名化措置によって、ライセンス型ミクロデータの安全性が担保されている。

- わが国において匿名データの作成・提供を検討する場合に、イギリスの事例は参考になるのではないか。

→イギリスにおける政府統計ミクロデータの作成状況については、今後も注視する必要があると考える。

主要参考文献

De Kort, S., and Wathan, J.(2009) “Guide to Imputation and Perturbation in the Samples of Anonymised Records”.

<http://www.ccsr.ac.uk/sars/resources/imputation.doc>.

藤野友和・垂水共之(2003)「PRAMの理論とその実用上の諸問題」『統計数理』第51巻第2号, 321～335頁

GSS Disclosure Control Policy for Microdata Produced from Social Surveys

<http://www.ons.gov.uk/ons/guide-method/best-practice/disclosure-control-policy-for-social-survey-microdata/index.html>.

伊藤伸介(2005)「イギリスにおけるミクロ統計データの提供とデータアーカイブの現状」『総務省統計局委託研究2004年度 統計データアーカイブに関する調査研究報告書』, 32～65頁

伊藤伸介(2011)「わが国におけるミ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ミクロデータを例に—」明海大学『経済学論集』第23巻第3号, 36～54頁

伊藤伸介(2012)「政府統計ミ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』第24巻第3号, 1～14頁

森 博美(2000)「イギリスにおけるミクロデータの提供」松田芳郎・濱砂敬郎・森博美編『講座ミクロ統計分析① 統計調査制度とミクロ統計の開示』日本評論社, 48～83頁

森 博美(2003)「イギリスにおけるデータアーカイブの現状—政府統計の二次的利用システム—」『平成14年度 国の統計調査に係るデータ・アーカイブに関する研究報告書』, 22～31頁

Shlomo, N.(2007) “Statistical Disclosure Control Methods for Census Frequency Tables”, *S3RI Methodology Working Papers M07/04*, pp.1-40.

<http://eprints.soton.ac.uk/44610/1/44610-01.pdf>.

Tranmer, M., Pickles, A., Fieldhouse, E., Elliot, M., Dale, A., Brown, M.(2005) “The Case for Small Area Microdata”, *Journal of Royal Statistical Society A*, Vol.168, pp.29-49.