

# 公的統計における行政記録データの利活用について

——デンマーク，オランダとイギリスの現状——

伊藤伸介

1. はじめに
2. 諸外国におけるセンサス統計の類型化
3. デンマークにおける行政記録の利活用について
4. オランダにおける行政記録の活用の展開
5. イギリスにおける行政記録の活用状況— Beyond 2011について—
6. むすびにかえて

## 1. はじめに

本稿の目的は、諸外国の公的統計（政府統計）における行政記録データの利活用の状況を見ていくことによって、行政記録データにおける利活用の方向性を模索することである<sup>1)</sup>。近年、「統計改革」という言葉が、話題になっている。2015年10月16日に開催された平成27年第16回経済財政諮問会議では、国内総生産（GDP）の推計に用いられる経済統計の精度に関して問題提起が出され、GDP統計の精度の改善を目指した経済統計のさらなる整備の必要性について議論がなされた。その後、2015年11月4日と2016年3月24日の経済財政諮問会議における議論を経て、平成28年第22回経済財政諮問会議（2016年12月21日開催）では、GDP統計の精度改善のための基礎統計の整備に加えて、欧米諸国で進められてきたEBPM（=Evidence Based Policy Making, 客観的な事実に基づい

---

1) 本稿は、2016年度統計関連学会連合大会における企画セッション「公的統計における新たな地平の拡大—ビッグデータと行政記録の可能性」(2016年9月5日、於 金沢大学)における発表内容に基づいている。本稿の作成においては、筆者が、イギリス国家統計局(2016年8月22日)、オランダ統計局(2016年8月24日)とデンマーク統計局(2016年8月30日)を対象に行ったヒアリング調査の内容と資料、およびオランダ統計局で開催された人口レジスターに関する講習会(2016年8月25~26日)での資料が用いられた。行政記録データの利活用に関して貴重な情報を提供していただいた、イギリス国家統計局のLouisa Blackwell氏、Neil Hopper氏、Nicola Rogers氏とPeter Jones氏、オランダ統計局のEric Nordholt氏、Peter-Paul de Wolf氏、Kees Prins氏、Maarten Emos氏とPaul Vander Laan氏、デンマーク統計局のIvan Thaulow氏、Jørn K. Petersen氏、Henrik Bang氏、Amy Frølander氏とHenning Chrishansen氏に記して感謝を申し上げる次第である。

て政策立案を行うこと)をわが国でも展開するために、公的統計のデータに加えて、民間のビックデータや行政記録データの利用可能性が指摘された。これらを踏まえる形で、2017年に統計改革推進会議(議長 菅義偉内閣官房長官)が設置された。

統計改革推進会議は、「政府全体における EBPM(証拠に基づく政策立案)の定着、国民のニーズへの対応等の統計行政部門を超えた見地から推進する」ことを目的として開催され、2017年5月には、『統計改革推進会議最終取りまとめ(以下『報告書』と呼称)』が発表された。『報告書』は、EBPMの推進体制の構築のために、「統計を始めとする各種データ」の整備・改善が必要であることを明記している。この「統計を始めとする各種データ」には、「統計、統計マイクロデータ及び統計的な利活用を行うために用いられる行政記録情報」が含まれている。こうした統計改革推進会議の動きを見ると、「統計改革」という形で、公的統計データの作成・提供や行政記録データの利活用のあり方への注目が高まっていることがわかる。

ところで、行政記録情報等の利活用の推進は、第Ⅱ期「公的統計の整備に関する基本的な計画」(平成26年3月25日閣議決定)においても明記されていた。その意味では、行政記録データのさらなる利活用は、公的統計の分野における政策的な課題の1つとして、これまでも議論されてきた。その一方で、個人情報の分野では、民間企業や行政機関が保有する個人情報に関して利活用と保護の両面からさらなる法整備を行うために、2015年8月に、改正個人情報保護法が成立しただけでなく(2017年5月に全面施行)、2016年5月に行政機関個人情報保護法・独立行政法人等個人情報保護法の改正も行われた。こうした個人情報の取り扱いに対する法制度面の変化は、わが国の公的統計の分野における行政記録データの利活用をめぐる議論に影響を与える可能性があるだろう。

公的統計の作成を指向した行政記録データの活用可能性は、これまでも多くの先行研究によって指摘されてきた。例えば、森(2006)は、方法的な視点さらには法制度的な観点から、わが国における行政記録データの公的統計への活用に向けた課題を指摘している。また、舟岡(2008)は、わが国の統計改革との関連で、諸外国における行政記録データの活用状況を法制度的な側面から明らかにした。さらに、濱砂(2008)は、ドイツ連邦統計局において検討が進められてきた統計登録簿型人口センサスの方法的な可能性を洞察した。

他方で、諸外国では、近年、学術研究目的のために行政記録データの利用や公的統計マイクロデータとリンクされた行政記録データの利用が広範に展開されてきた。例えば、イギリスでは、Administrative Data Research Network(ADRN)と呼ばれるプロジェクトが広範に展開されており、学術研究を通じた行政記録データの提供や公的統計のマイクロデータと行政記録データのリンクが行われている(伊藤(2016))。

UNECE主催の人口・住宅センサス専門家会議(2015年9月~10月、Geneva)では、行政記録に基づく効率的な統計の作成も主要な議題の1つとして討議され、公的統計の作成を指向した行政記録データの活用が国際的な注目を集めた。例えば、UNECE主催の移民統計に関するワークセッ

ション(2016年5月, Geneva)では、オーストラリアやカナダにおける行政記録の活用事例について報告が行われた。また、イギリスでは、パスポートの情報を利用して外国人の移出入の状況を捕捉するための調査研究プロジェクト(Semaphore Research)が存在する。このように、行政記録の利活用をめぐる状況は、諸外国では多様な様相を呈している。

そこで、本稿では、主として、行政記録データの提供の先進国と言えるデンマークとオランダ、さらには行政記録データの活用可能性に関する研究を進めているイギリスを対象に、諸外国の公的統計に対する行政記録データの利活用の現状を明らかにすることによって、行政記録データにおける利活用の方向性を洞察する。

## 2. 諸外国におけるセンサス統計の類型化

行政記録に基づく統計作成をめぐる状況は、諸外国では多様な様相を呈している。1970年代以降、多くのヨーロッパ諸国では、市民のプライバシー意識の顕在化に伴い、調査票ベースの人口センサスの実施が困難な状況にあった。そこで、統計作成部局は行政記録に基づくセンサスの実施可能性を模索した(工藤(1995a))。それによって、欧米諸国においては、①行政記録に基づいた統計作成システムの確立、②調査票ベースの統計作成システムの採用、③行政記録と調査票の両方を用いた統計作成システムの展開、の3つの方向で統計作成が行われているのが現状である。

諸外国において人口センサスの作成においては、以下の3つのタイプに類型化することが可能である(Statistics Netherlands(2014))。第1のタイプは、レジスター型センサス(register-based census)と呼ばれるセンサス統計である。これについては、行政記録データに基づいた統計作成が行われる。現在、レジスター型センサスを採用している国については、当初は統計調査に基づく統計の作成を行ってきたが、行政記録データに基づく統計作成の可能性を検討する中で、段階的にレジスター型センサスに移行し、現在、レジスターベースの統計作成システムが確立している。こういった統計作成を行っている国としては、デンマーク、フィンランド、スウェーデン、ノルウェーといった多くの北欧諸国だけでなく、オーストリア、スロベニア、ベルギー、アイスランド、オランダといった国が該当する。

第2のタイプは、混合型センサス(combined census)である。これは、調査票ベースとレジスターベースの両方によるセンサス統計の作成を包含している。統計作成にあたって、調査票と行政記録のそれぞれがどの程度の割合で活用されているかについては、国によって異なると考えられる。このような統計作成を行っている国としては、エストニア、ラトビア、リヒテンシュタイン、リトアニア、ドイツ、イスラエル、ポーランド、スペイン、スイス、トルコ、チェコといった国を指摘することができる。これらの国については、現段階では混合型センサスを実施しているが、将来的には、レジスター型センサスに移行する可能性がある国も存在する(例えば、チェコの事例)。

第3のタイプは、伝統型センサス (traditional census) である。これは、統計作成部局が実施してきた統計調査から得られた調査票ベースの人口センサスの作成を指向している。これについては、アメリカ、イギリス、カナダ、オーストラリア、ニュージーランド等の多くの国々が、それに該当する<sup>2)</sup>。その一方で、伝統型センサスを採用している国において、行政記録データに基づく統計作成の可能性を模索していることは注目すべき点であると言えよう。例えば、カナダにおいては、カナダ統計局が、税務データ等の行政記録を用いて人口センサスの作成可能性を追究している (Lebel and Denis (2016))。

次節以降では、世界で初めてレジスター型センサスに移行したデンマーク、virtual census という統計作成システムを採用しているオランダ、さらに、伝統型センサスを実施しながらも行政記録データの活用に向けた検討を進めているイギリスの事例を紹介することにしたい。

### 3. デンマークにおける行政記録の利活用について<sup>3)</sup>

本節では、デンマークにおける公的統計への行政記録データの活用の現状および行政記録データの提供状況について見ていくことにする。デンマークは、最初に調査票ベースのセンサスからレジスター型センサスに移行した国として知られている。行政機関は国民の個人情報行政記録という形で管理するが、国民に対して行政サービスの提供を行うという国・行政機関と国民との間の信頼関係が、こうしたレジスター型センサスの成立を可能にしている (United Nations Economic Commission for Europe (2007))。デンマークでは、レジスター化を進めるために、1968年に中央人口レジスター (Central Population Register=CPR) が内務省によって創設された。最後の調査票ベースのセンサス (伝統型センサス) が実施された1970年には、税務レジスター (Tax Register) の整備が行われた。そして、1974年には教育レジスター (Education Register) が、1975年に企業・事業所レジスター (Register on Enterprises and Establishments=CBR) が、さらに、1977年には建物・住居レジスター (Buildings and Dwellings Register=BDR) がそれぞれ整備されている。これらのレジスターの整備が進められたことは、デンマークにおいてレジスター型センサスに移行するための準備が段階的に進められてきたことを示唆している。さらに、1979年には、CPRにおいて父親と母親の個人番号が導入された。これによって、レジスターにおける家族構成員の連結が可能になった。こうした状況を受けて、1981年に、第1回のレジスター型センサスに基づく統計の作成が行われた。それ以降、デンマークでは、レジスターベースのセンサス統計の作成が展開される。1990年

---

2) 伝統型センサスについては、フランスのローリングセンサスを含める場合があるが、ローリングセンサスに関しては、別の類型として位置付けることもある。

3) 本節は、工藤 (1990, 1995b) を考察の素材にしつつ、2016年8月30日に筆者がデンマーク統計局で実施したインタビューの内容に基づいている。

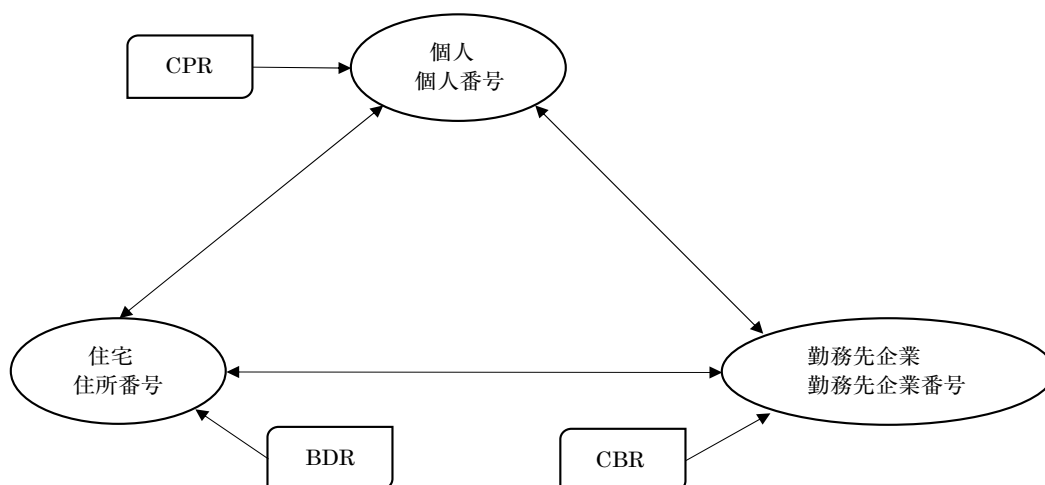
～1991年には、CPRにおいてオンラインによるデータの更新が導入されたことによって、レジスター情報におけるタイムリーな更新が可能になった。さらに、2007年には、デンマーク統計局において人口統計データベース(Person Statistics Databases=PSD)が構築された。

デンマークにおいては、中央人口レジスター(CPR)が、個々人の人口社会的属性に関する行政記録情報の一元的な管理を行っている。CPRでは、個人番号を管理しており、個人や世帯の人口社会的属性として、名前、年齢、性別、結婚状態、市民権、出生地と住所が把握されるだけでなく、家族情報(母親、父親と配偶者のCPR)もレジスターに登録されている。

図1は、デンマークの情報システムにおけるレジスターの基本モデルを示している。CPRでは個人番号(Personal number)が個人の識別子として登録されている。それに対して、建物・住居レジスター(BDR)に住所番号(Address number)が登録されていることから、住所番号は、CPRとは異なるレジスターのシステムで管理されている。さらに、勤務先企業についても、企業・事業所レジスター(CBR)による勤務先企業番号(CBR Workplace number)の管理が行われている。これらの個人番号、住所番号と勤務先企業番号は、相互に連結されている。なお、世帯に関する識別子は存在しないが、CPRに基づく個人情報および住所情報によって、世帯の把握も可能になっている。

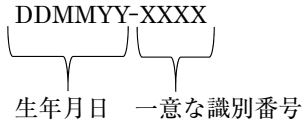
個人番号はCPR(Civil Personal Registration)番号とも呼ばれており、デンマークに居住する個人に割り当てられるユニークなIDを表している。具体的には、生年月日、性別、年齢に関する情報が個人番号に含まれている。個人番号は10桁の数字から成っている。最初の6桁が生年月日を表しており、残りの4桁が一意的な識別番号を表している。

図1 デンマークの情報システムにおけるレジスターの基本モデル



注) デンマーク統計局におけるヒアリング調査での内容と資料に基づいて作成

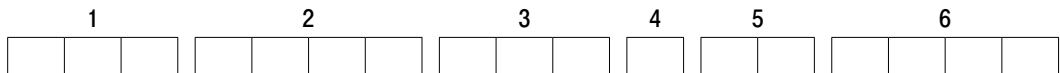
## 個人番号のイメージ



最後の1桁は、性別に関する情報であって、0, 2, 4, 6あるいは8であれば、対象となる個人が女性であることを表すが、1, 3, 5, 7かあるいは9であれば、その個人は男性であることを示している。なお、デンマークに3か月以上居住している個人はすべて、CPR番号を取得する必要がある。

住所番号のイメージは、以下のように示される。

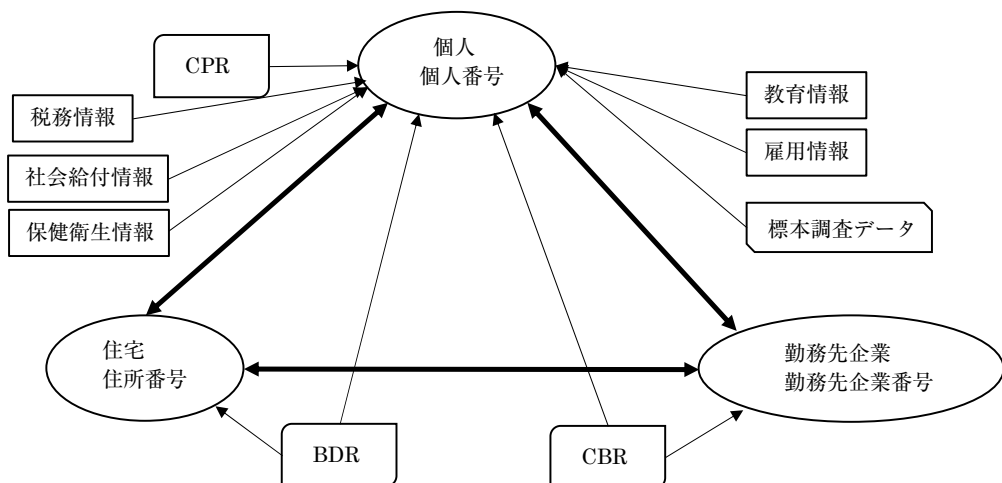
## 住所番号のイメージ



住所番号は、17桁のコードから構成される。具体的には、1は地方自治体コード、2は道路コード、3は番地、4は住宅の名称、5は階数、6はドア番号である。これらの番号によって、住所を一意に識別することが可能になる。

図2は、デンマークにおける統計情報システムを示したものである。デンマークの場合、税務情報、社会給付状況、保健衛生情報、教育情報、雇用情報といった社会人口的な属性情報が行政記録

図2 デンマークにおける統計情報システム

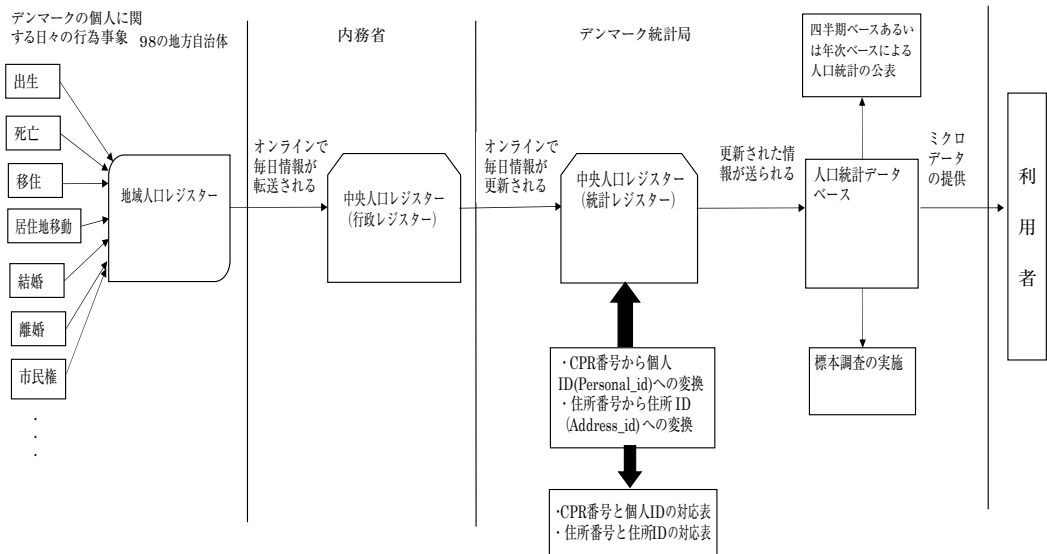


注) デンマーク統計局におけるヒアリング調査での内容と資料に基づいて作成

によって把握することができるだけでなく、個人番号によって連結することが可能になっている。また、先述のとおり、住所番号と勤務先企業番号は、個人番号と一意に対応可能になっている。したがって、基本的には、統計作成において必要な属性情報については行政記録から獲得することが可能になっている。一方、行政記録から入手できない属性情報については、標本調査が実施される<sup>4)</sup>。なお、標本調査のマイクロデータは、個人番号を通じて連結することが可能になっている。このように労働供給側である個々人の就業属性と労働需要側である個人が勤務する企業の特性も連結されていることから、例えば、労働需要と労働供給の両面からの労働市場分析を行うことも可能なデータ構造が備わっている。

つぎに、2007年に構築された人口統計データベースについて見ていきたい。図3は人口統計データベースによる統計作成の概略図である。出生、死亡、移住、居住地移動、結婚、離婚、市民としての生活状況といったデンマークの個人に関する日々の行為事象の結果は、98の地方自治体における地域人口レジスターにおいて逐次反映される。地域人口レジスターに登録されている地域レベルの個々人に関する行政記録情報は、内務省が管理している中央人口レジスターにオンラインで日々転送される。この中央人口レジスターは、行政目的のために収集された「行政レジスター（administrative register）」である。つぎに、内務省の中央人口レジスターに転送された情報は、デンマーク統計局が保管している中央人口レジスターにオンラインで送られることによって、レジスターに

図3 デンマークにおける人口統計データベースに基づく統計作成の概略図



注) デンマーク統計局におけるヒアリング調査での内容と資料に基づいて作成

4) デンマーク統計局の担当者によれば、Eurostat（欧州統計局）の要請で、国際比較を可能にする労働力調査をデンマーク統計局も実施しているとのことであった。

登録される情報が毎日更新されている。このデンマーク統計局に備わっている中央人口レジスターは、統計目的のために管理・更新される「統計レジスター (statistical register)」である。デンマーク統計局が管理する統計レジスターは、内務省が保有する行政レジスターと基本的に同一のデータ構造を有している。しかしながら、デンマーク統計局が保管する統計レジスターの場合、CPR 番号と個人 ID (Person\_id) との対応表、住所番号と住所 ID (Address\_id) との対応表にしたがって、CPR 番号から個人 ID、住所番号から住所 ID への変換がなされた上で、統計レジスターに個人 ID と住所 ID が備えられている。このことは、地方自治体が行政サービスを行うために収集・登録された行政レジスターに含まれる個体情報が、統計レジスターにおいては、個人番号や住所番号といった個体識別番号を削除した上で、統計作成のための統計単位情報に転換されていることを示している<sup>5)</sup>。なお、この統計レジスターには、デンマーク統計局の職員はアクセスすることができるが、CPR 番号と個人 ID との対応表および住所番号と住所 ID との対応表はセキュアな環境で管理されており、一部のごく限られた職員のみがアクセス可能になっている。デンマーク統計局内部において、人口統計データベース (PSD) を編成するために、中央人口レジスターから更新された情報が逐次送られている<sup>6)</sup>。この人口統計データベースを集計することによって、四半期ベースや年次ベースによる人口統計が公表されている。また、人口統計データベースは、標本調査を実施するためのサンプリングフレームとしても用いられている。さらに、人口統計データベースから抽出されたマイクロデータが利用者に対して提供されている。

デンマークにおけるマイクロデータの提供状況は、以下のとおりである。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ことが特徴的である。2001年より、リモートアクセスのサービスが開始されている。一方、デンマークでは、他の欧米諸国のように、匿名化マイクロデータ (anonymized microdata) やオンデマンド集計システムといったマイクロデータの提供に関する他のチャンネルは存在しない。

デンマークにおいては、個票データのアクセスが認証された大学等の機関においてのみ個票データの提供が可能である。これについては、2016年8月時点で、400以上の認証された機関が存在する。また、1100以上の研究プロジェクトによって個票データの利用が行われており、1800人以上の研究者が個票データにアクセスしていることが確認されている (2016年8月時点)<sup>7)</sup>。個票データの利用

5) 行政記録情報における個人番号や住所番号といった個体識別情報をどのように考えるかについては、浜砂 (2010) が「行政個別情報と統計単位情報の情報的な相違 (例えば、前者の非単位性や非同時性と後者の単位性と同時性) や法制度的な条件の相違 (例えば、前者の個別的・非匿名的利用と後者の匿名的集計の利用)」を指摘している (浜砂 (2010, 32頁))。

6) 人口統計データベースの管理に関する興味深い点として、デンマーク統計局は、人口統計データベースの修正を行わないことが指摘される。もし、個人の登録情報に何らかのテクニカルな誤りが見つかった場合には、中央人口レジスター (CPR) において修正がなされる。この CPR から更新されたデータが、継続的に PSD に入力されている。



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である。さらに、人口社会、保健衛生、所得、教育、労働市場、企業経済等の様々な分野の個票データが利用可能である<sup>8)</sup>。

デンマークでは、「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である。一方で、マイクロデータに含まれる個人情報秘密保護の観点から、個人データの処理に関する法律 (Danish Act on Processing in Personal Data)、1985年行政機関法 (Public Administration Act 1985)、デンマーク刑法 (Danish Penal Code) といった法律に基づいて、マイクロデータが提供されている。

デンマークにおいては、リモートアクセスによる個票データの分析結果 (集計表、回帰分析等) に関する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、職員が事後的にマニュアルでチェックを行う。したがって、分析結果のチェックにおいて自動化は行われていない。職員が分析結果をチェックする上では、閾値ルール (thresholding rule) のような基準は採用されている。また、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いるが、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。なお、デンマークにおいても、分析結果のチェックに関しては、基本的には、統計作成部局と研究者との信頼関係が重視されている。

#### 4. オランダにおける行政記録の活用の展開<sup>9)</sup>

本節では、virtual census という統計作成システムを採用しているオランダを例に、オランダにおける行政記録の活用の展開について見ていくことにしたい。森 (2009b) によれば、オランダにおける virtual census は、以下のように展開されている。1971年に、オランダ統計局によって、最後の伝統型センサスが実施される。その後、調査における非回答の存在と調査にかかるコストの削減の必要性によって、伝統型センサスの実施が困難となる。そのために、1981年と1991年において

---

7) デンマーク以外の研究者の場合、デンマークに滞在せずに、個票データのアクセスが認証された大学等の機関を通じて、個票データにリモートアクセスを行うことができる。この場合には、認証された大学等の機関がその研究者が学術目的のために安全に個票データを利用することに関して保証することが、利用の条件となる。

8) デンマーク統計局の担当者によれば、同一の研究者が、関連する新規のプロジェクトにおいて個票データを利用するために、個票データの利用期間を更新し、長期にわたって個票データを利用しているケースも存在するとのことであった。

9) 本節は、森 (2009a, 2009b) を参考にしながら、2016年8月24日に筆者がオランダ統計局の担当者に行ったヒアリング調査、および2016年8月25日~26日に筆者が参加した人口レジスターに関する講習会 (於 オランダ統計局) の資料に基づいている。

は、人口レジスターと調査データに基づく限定的な virtual census が実施された。

その後、1990年代においてレジスターが整備されることによって、2001年と2011年において、社会統計データベース（Social Statistics Databases=SSD）（森（2009a））に基づく完全な virtual census が実施された。これは、人口レジスターを中心とするレジスターベースのセンサスであるが、統計は、オランダ統計局外部の登録情報（レジスター）に基づいて作成される。登録データに存在しない場合にのみ、オランダ統計局によって、統計調査を実施することが認められている。こうしたことから、オランダで統計を作成する上では、行政記録データの整備が求められる。

オランダ統計局において統計を作成する上で用いられる基本レジスター（basic registers）は、以下のレジスターから構成されている（2016年8月時点）。

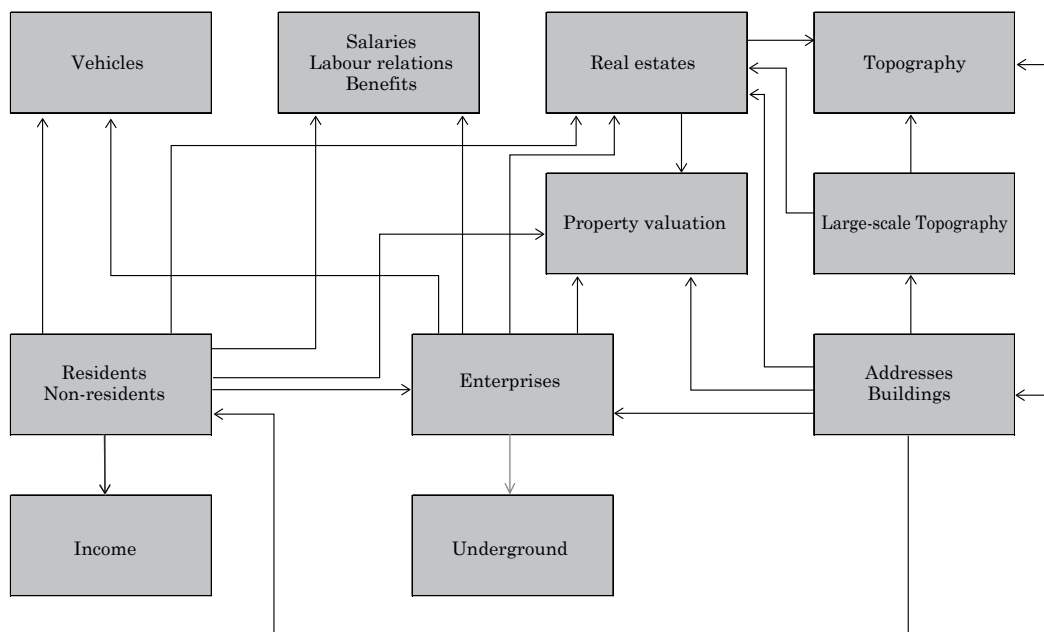
- ・ 人口情報（居住者 / 非居住者）（内務省（Minister of the Interior））
- ・ 住所・建物情報（基盤・環境省（Minister of Infrastructure and Environment））
- ・ 企業情報（経済省（Minister of Economic Affairs））
- ・ 不動産情報（基盤・環境省、金融省（Minister of Finance））
- ・ 地形情報（topography）（基盤・環境省）
- ・ 自動車情報（基盤・環境省）
- ・ 課税所得（金融省）
- ・ 労働情報（社会省（Minister of Social Affairs））
- ・ 地下情報（subsoil）（基盤・環境省）

この中で、人口情報（居住者情報）、住所・建物情報、不動産（価格）情報と地形情報は、地方自治体によって保有されている。

図4は、基本レジスターの体系図を示したものである。人口情報（居住者情報）レジスター、住所・建物情報レジスター、および企業情報レジスターを起点として、他のレジスターが連結されていることがわかる。また、人口情報、住所・建物情報と企業情報も、互いに連結されている。人口情報（居住者情報）の基本レジスター（BRP=Basic Registration on Population）は、地方自治体が管理する人口情報に関するデジタル化されたレジスターシステムであって、すべての人口動態事象が登録される。各居住者におけるすべての個人データ（名前、性別、生年月日、出生国、居住地等）が、個人レコードの電子媒体という形で保管されている。地方自治体が管理する居住者の登録情報のコピーが、内務省によって保管される中央レジスター（GBA-V）に送付される。なお、個人レコードに含まれる情報は、以下のとおりである。

1. 居住者情報（名前、性別、生年月日、個人識別番号（PIN））

図4 基本レジスターの体系図



注) オランダ統計局における人口レジスターに関する講演会の資料に基づいている

2. (母) 親の情報
3. (父) 親の情報
4. 市民権の状況
5. 配偶者（登録されたパートナー）の情報
6. 死亡の有無
7. 住所
8. 子供の情報
9. 法律上の地位（オランダ人以外の市民のみ）
10. パスポート／ID カード情報
11. 選挙の参加資格

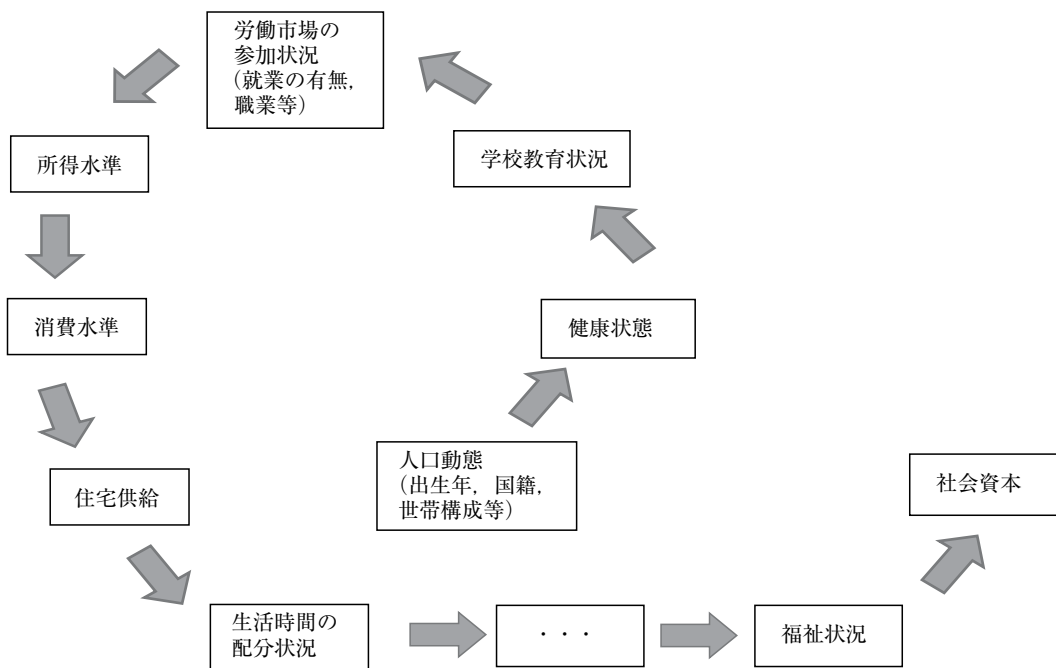
オランダでは、現在、2011年センサスが、レジスター型センサスが実施された最新年次となっている。2011年センサスのデータ源は、地方自治体で整備されている行政記録データであるが、その主要な行政記録データは、人口レジスター（population register）であって、それは、センサス統計の作成において基本となる行政情報である。また、すべての雇用者が登録対象になっている職業ファイル（Jobs file）やすべての自営業者が対象になっている自営業者ファイル（Self-employed file）、社会保障給付情報、年金・生命保険給付情報、住宅レジスターも、センサス統計の作成のため

めに用いられる。これらは、後述する SSD のデータ源にもなっている。センサス統計の作成において必要な学歴と職業については、統計作成の時点では、それらに関するレジスターが十分に整備されていなかったことから、標本調査として実施された労働力調査のデータが用いられている。なお、デンマークと同様、オランダにおいても世帯レジスターは存在しない。

ところで、SSD とは、人口レジスターを中心に編成されたデータベースで、個人・世帯属性、就業属性、社会保障給付状況、所得情報、学校教育状況、健康状態、住宅属性といった様々な情報が連結されるだけでなく、労働力調査といった標本調査データも連結されている。図 5 は、SSD におけるライフサイクルモデルの模式図を示したものであるが、SSD の場合、原データとなる行政記録データに含まれる個人・家族の属性が、地方自治体レベルで逐次更新されることによって、出生年、国籍、さらに世帯構成といった人口社会的特性、健康状態、学校教育状況、就業の有無や職業といった労働市場の参加状況、所得水準、消費水準、住宅供給、生活時間の配分状況といった社会人口事象や経済事象が逐次的に連結され、更新される縦断的な構造を備えている。その意味では、SSD は、マイクロシミュレーションモデルによって行われる個別主体群を含むデータベースの更新と同様のデータ構造を備えていると考えることができる（伊藤（2002, 2013））。

行政記録データ（レジスター）間のリンケージにおいては、一意な識別番号（リンケージキー）として、市民サービス番号（Civil service number）（2007年以前は、社会保障番号（Social security

図 5 SSD におけるライフサイクルモデルの模式図



注) オランダ統計局における人口レジスターに関する講演会の資料に基づいて作成

and Fiscal number (SoFi) が用いられていた) (森 (2009a)) が使用されている。一方、標本調査データとのリンケージにおいては、性別、生年月日、住所(郵便番号と住宅コード)が用いられる。いずれもマッチング率はほぼ100%である。なお、リンケージキーは、SSDにおいては、RIN-Person (RINは、レコード識別番号(Record Identification Number)の略称)という識別番号に置き換えられる。

オランダについても、行政記録データが公的統計を作成する上でのデータ源になっていることから、行政記録データの二次利用の観点から、マイクロデータの提供状況についても見ていくことにしたい。

オランダの場合、匿名化マイクロデータについては、1971年、2001年と2011年の人口センサスの匿名化マイクロデータが1%の標本データとして利用可能である。主として非攪乱的な手法(non-perturbative method)が適用されているが、マイクロデータのデータ特性によっては、攪乱的な手法(パータベーション, perturbation)も適用される。具体的には、匿名化マイクロデータの作成において、PRAM (=Post Randomization Methods)による秘匿処理が用いられる場合がある。また、匿名化手法の1つとして、地域、性別、年齢のような変数を用いた低次元のクロス表をもとに、特殊な一意(special unique)に対する処理がなされている。一方、オランダ統計局は、 $\mu$ -Argusと呼ばれるマイクロデータの秘匿処理用のソフトウェアを用いて、public use microdata filesの作成を行っている(Nordholt (2013))。

個票データについては、オンサイト施設(オランダ統計局内部に2か所)やリモートアクセスによる利用が可能である。興味深いのは、先述のSSDについても個票データとして利用することができることである。ただし、分析に必要な変数のみがアクセス可能となっている。個票データを分析した後の集計表や帰帰分析の結果等のチェックに関しては、オランダにおいても、デンマークと同様に、マニュアルによるチェックがオランダ統計局の担当職員によって行われている。

他方、オランダ統計局では、StatLineと呼ばれるオンデマンド型の集計システムが開発されている。StatLineの場合、オンデマンド型集計システムの保管用データとして、データキューブ(hyper cube, 超高次元クロス集計表(伊藤(2009))がオランダ統計局によって編成されている。このStatLineの保管用データとしてのデータキューブは、4~5変数からなる集計表の集合体であって、集計表に用いられる変数の中に、センシティブな変数は含まれていないことが知られている。

## 5. イギリスにおける行政記録の活用状況—Beyond 2011について—

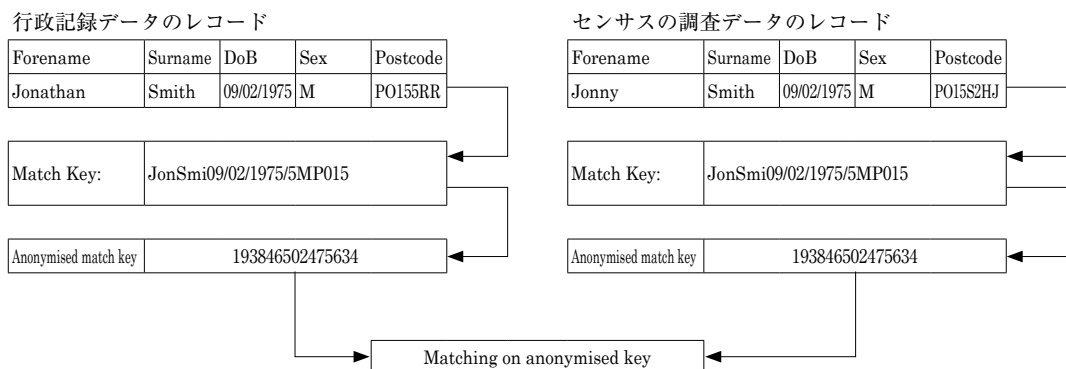
最後に、イギリスにおける行政記録の活用状況の現状についても簡単に見ていくことにしたい。伝統型センサスの実施国であったイギリスにおいても、行政記録データのさらなる活用という観点から、イギリス国家統計局において、「Beyond 2011」というプログラムが2011年4月に創設され

た。2014年3月に、イギリス国家統計局長は、今後の人口統計の作成および2021センサスにおける勧告を行った。それは、2021年センサスにおける統計作成のさらなる展開およびセンサス間の年次ベースの統計の改善を目指して、行政記録データの利活用を推奨するものであった。

こうした状況を踏まえて、Beyond 2011は、2011年～2015年に実施された。それは、人口・住宅統計の作成に関する様々なアプローチを検証するためのプロジェクトであり、人口推計の過小ないしは過大把握の評価を行うために行政記録が用いられた。患者レジスター（NHS General Practitioner Patient Register）、学生レコード（the student Higher Education Statistics Agency Student Record dataset）、出生レジスター（the Live Births Register）、死亡レジスター（the Deaths Register）、選挙レジスター（Electoral Registers）等、レジスターの中でも捕捉の程度が高い行政記録データが人口推計値の精度を検証するために用いられ、それに関して各種の行政記録間のレコードリンケージの方法が検討された（Blackwell et al. (2015)）。さらに、人口センサスのデータとの比較検証によって行政記録データの精度の評価を行うために、統計的人口データベース（Statistical Population Database=SPD）が構築された。Beyond 2011では、行政記録からSPDに関して様々なテストデータを作成し、その精度に関する検証を行っている（Office for National Statistics (2013a)）。

Beyond 2011においては、SPDを構築にあたって、患者レジスターや学生レコードといった行政記録データに対して、確率的マッチング、スコアに基づくマッチングさらにはロジット回帰モデルに基づくマッチング等、様々なマッチング方法に関する研究が行われている（Office for National Statistics (2013b)）。イギリスの場合、デンマークやオランダとは異なり、リンケージを可能にするための共通のIDが個別レコードには存在しない。そこで、名前と住所を用いた確率的なリンケージが行われる。図6は、仮名化による行政記録データのレコードとセンサス調査データのレコードとのマッチングに関する概略図を示したものである。行政記録データにあるJonathanという名前とセンサス調査データにおけるJonnyという名前が異なることから、生年月日、性別、さ

図6 仮名化によるマッチングの例



出所) Office for National Statistics (2013b, p.10)

らには地域区分が同じであっても、名寄せによるマッチングは成功しない。そこで、名前と住所をそのまま用いてリンケージするのではなく、IDを付与するためにハッシュ化が行われている(仮名化(pseudomization))。図6では、「JonSmi09/02/1975/5 MP015」というマッチングキーに変換され、さらに「193846502475634」という匿名化されたマッチングキーに変換された上で、リンケージが行われている(Office for National Statistics (2013b))。リンケージは第3者機関において行われており、統計局の職員は、直接IDに触れることはできないことから、IDに関する秘密保護が確保されている。

なお、現在、Beyond 2011の後継プロジェクトとして、2015年よりセンサス転換計画(Census Transformation Programme=CTP)が進められており、2021年センサスに向けて、行政記録データの活用可能性が模索されている(Office for National Statistics (2016))。

## 6. むすびにかえて

本稿では、デンマーク、オランダとイギリスを例に、公的統計における行政記録データの利活用の現状を明らかにした。行政記録データにおける連結を可能にするためには、IDの整備と管理が必要である。デンマークやオランダとは異なり、行政記録データにおいて連結するための共通のIDが存在しないイギリスでは、仮名化を行うことによって、IDを付与した上で、行政記録データのリンケージに関する実験が行われている。

本稿で見てきたレジスター型センサスについては、その実施において利点と欠点が存在する。レジスター型センサスの利点については、①センサス統計の作成に伴う費用の削減、②非回答の回避、③更新された登録情報に基づく早期の公表可能性を指摘することができる。他方で、レジスター型センサスの欠点としては、①レジスターにおける概念と統計における集計事項の概念の相違について調整が求められること、②地方自治体レベルだけでなく、統計作成部局においても登録情報の更新が必要なこと、③統計作成におけるインフラの整備(行政記録を更新するためのオンライン化等)が求められることが知られている。

レジスター型センサスにおいては、行政記録データを更新することによって、個体レコードの情報が絶えず更新されるシステムが採用されている。伝統型センサスにおいては、それはある時点のスナップショットであるが、レジスター型センサスにおいては、個体レコードそれ自体が縦断的な構造をそなえていることがわかる。それによって、人口の将来予測が可能になる。このことは、レジスター型センサスがマイクロシミュレーションへの展開可能性を有していることを示唆している。

ところで、本稿の最初に述べたように、わが国では、公的統計のデータに加えて、民間のビックデータの活用可能性が注目されている。海外の統計作成部局のビックデータの活用に関する状況について見ていくと、オランダ統計局では、ビックデータを用いた統計作成に関する研究が現在行わ

れており、そのために Innovation Lab という施設が、2012年5月に統計局内部に設置されている。例えば、ソーシャルメディア（Twitter等）、道路センサーのデータ（Traffic loop data）といったビックデータの利用可能性が追究されている。さらに、オランダ統計局の場合、将来的には、ビックデータを SSD のデータ源に含めて、SSD を拡張する可能性が模索されている。また、イギリス国家統計局においては、センサス転換プロジェクトにおいて、ビックデータとしての携帯電話情報の利用可能性が追究されている。なお、デンマーク統計局では、ビックデータの可能性について関心を持っているものの、ビックデータを用いた統計作成の可能性に関する研究は行われていない。

行政記録データや民間のビックデータも含め、諸外国では、公的統計を作成するためのデータ収集の方法は、①費用の効率性、②回答者の負担、③統計の品質のバランスを勘案した上で決定される。海外の統計作成部局においても、伝統型センサス、混合型センサスとレジスター型センサスのいずれかが指向されるにあたって、データ収集に関するバランスが作用しているものと考えられる。その意味では、わが国においても、個人情報の取り扱いに対する文化的環境や国民性の違いさらにはその法制度上の位置づけも勘案した上で、公的統計の作成における行政記録データやビックデータの利活用のあり方が、今後も議論されていくだろう。

付記：本稿は、「2016年度中央大学特定課題研究費」における成果の一部を発表したものである。

#### 参考文献

- 伊藤伸介（2002）「アメリカにおけるマイクロ社会モデルの体系化の試み—オーカットの社会人口モデルと所得移転モデル—」、『統計学』第83号，11-31頁
- 伊藤伸介（2009）「匿名化技法としてのマイクロアグリゲーションについて」熊本学園大学『経済論集』第15巻第3・4号合併号，197-232頁
- 伊藤伸介（2013）「マイクロシミュレーションモデルの方法的展開—アメリカのマイクロ社会分析モデルを例に一—」，明海大学『経済学論集』Vol. 26, No. 1, 1-18頁
- 伊藤伸介（2016）「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—イギリスを例に一—」、『経済学論纂（中央大学）』第56巻第5・6合併号，1-19頁
- 工藤弘安（1990）「統計調査における情報提供（Ⅱ）—事例研究：デンマークその1—」『成城大学経済研究』108号，45-60頁
- 工藤弘安（1995a）「レジスター・ベースの人口・住宅センサス」『成城大学経済研究』127号，71-87頁
- 工藤弘安（1995b）「統計調査における情報提供（Ⅲ）—事例研究：デンマークその2—」『成城大学経済研究』128号，98-79頁
- 濱砂敬郎（2008）「現代センサス革命の一断面—ドイツの2011年統計登録簿型人口センサスについて—」熊本学園大学『経済論集』第15巻第3・4号合併号，35-58頁
- 浜砂敬郎（2010）「2007年統計法といわゆる「基本計画」について」『経済学研究』（九州大学）第77巻第1号，27-44頁
- 舟岡史雄（2008）「各国の統計法制度とわが国の統計改革」国友直人・山本拓編『21世紀の統計科学Ⅰ 社会・経済の統計科学』東京大学出版会，147-170頁



- 森博美 (2006) 「諸外国における行政情報の統計利用の現状とわが国統計の課題」『経済志林』第73巻第3号, 817-869頁
- 森博美 (2009a) 「オランダの社会統計データベース SSD について」『経済志林』第76巻第4号, 5-28頁
- 森博美 (2009b) 「オランダの virtual census について」熊本学園大学『経済論集』第15巻第3・4号合併号, 35-58頁
- Blackwell, L., A. Charlesworth, N. L. Rogers (2015) “Linkage of Census and Administrative Data to Quality Assure the 2011 Census for England and Wales”, *Journal of Official Statistics*, Vol. 31, No. 3, pp. 453-473.
- Lebel, A. and J. Denis (2016) “Assessing the usability of a statistical population register for the Census of Population in Canada”, Paper Presented at the United Nations Economic Commission for Europe, Conference of European Statisticians, Group of Experts on Population and Housing Censuses, Geneva, pp.1-16.
- 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.
- Office for National Statistics (2013a) Beyond 2011: Producing Population Estimates Using Administrative Data: In Practice  
(<https://www.ons.gov.uk/census/censustransformationprogramme/beyond2011censustransformationprogramme/reportsandpublications>) [2017年8月1日アクセス]
- Office for National Statistics (2013b) Beyond 2011: Matching Anonymous Data  
(<https://www.ons.gov.uk/census/censustransformationprogramme/beyond2011censustransformationprogramme/reportsandpublications>) [2017年8月1日アクセス]
- Office for National Statistics (2016) 2021 Census Design Document  
(<https://www.ons.gov.uk/census/censustransformationprogramme>) [2017年8月1日アクセス]
- Statistics Netherlands (2014) *Dutch Census 2011: Analysis and Methodology*, Statistics Netherlands.
- United Nations Economic Commission for Europe (2007) *Register-Based Statistics in the Nordic Countries: Review of Best Practices with Focus on Population and Social Statistics*, United Nations.

(中央大学経済学部教授 博士(経済学))

