

S. ヴェイル(Steven Vale : 国連欧州経済委員会)編
著『政府統計のための行政的なデータソースと第2次
的なデータソースの利用 : 原則と実務にかんするハ
ンドブック』第1部

浜砂, 敬郎
九州大学 : 名誉教授

西村, 善博
大分大学経済学部 : 教授

<https://doi.org/10.15017/1446179>

出版情報 : 経済学研究. 80 (5/6), pp.211-249, 2014-03-31. 九州大学経済学会
バージョン :
権利関係 :

(翻訳資料)

S. ヴェイル (Steven Vale : 国連欧州経済委員会) 編著

『政府統計のための行政的なデータソースと第2次的なデータソースの利用
—原則と実務にかんするハンドブック—』第1部

浜砂 敬郎 西村 善博 分担・共訳

訳者序文

センサス革命という表現に形容されるような統計システムの改編が世界的に進展している。システム改編の基本的な方向は、統計作成過程における行政登録簿 (administrative register) の本格的な導入であって、政府統計全体を、基本的には行政登録簿から策定した統計登録簿にもとづいて作成するスカンディナヴィア諸国の統計システムを、中欧諸国、さらには多くの欧州の大陸諸国における統計システムによって、全体的にあるいは部分的に導入することが開始されている^(注 i)。

英米両国に目を向けると、イギリスでは、すでに企業統計登録簿が策定され、2021年人口・住宅センサスは、行政登録簿にもとづいて実施されることが決定している。またアメリカでは、経済センサスの小さくない部分が行政登録簿の情報によって補完されており、人口・住宅センサスもセンサスの調査対象者表である地理的住所システムが行政登録簿によって作成されている。

さらに、近年、わが国でも、国勢調査における未回答者を住民登録簿によって補記することが本格的に導入され、経済センサスの母集団リストが商業登記簿や税務データによって補完する作業が進められている。

各国の統計登録簿と政府省庁の行政登録簿が関連しているために、登録簿統計の研究は、各国の行政登録簿にかんする具体的な分析を欠くことはできない。そのために、本領域の研究では、国際的な動向を一般的に把握するとともに、各国の個性的な行政登録簿にかんする深い考察が必要である。ここに訳出するハンドブックは、言わば、その「中間」を行く実務的なテキストであって、国際連合欧州経済委員会が、世界各国の政府統計関係者にたいして、2001年から10回ほど開催してきた登録簿統計にかんする国際セミナーにおける講義と討論を基礎にしている。最近では、2011年6月6～9日にフランス Libourne 市の CEFIL-Centre de Formation de L'INSEE において開かれ、本書は同セミナー後に、セミナーの指導者である S. ヴェイル氏によって編成、国際連合欧州経済委員会より刊行され、2013年1月には、Wiki 版が Web サイトに掲載されている。フランス Libourne 市におけるセミナーの

(注 i) 浜砂敬郎「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合併号、2009年第1節参照

内容・状況の一端および雰囲気は、セミナー冒頭におけるヴェイル氏のガイダンスと講義において使用された、つぎのスライドから伺うことができる。

(スライド1) 統計目的のための行政的なデータソースの利用

スティーヴン ヴェイル (国際連合欧州経済委員会)

(スライド2: プログラム)

第1日目

09.00-10.00: はじめに一課程の概略—

10.00-12.30: 行政的なデータソース—定義、利点と精度の考察—

14.00-15.30: 行政的なデータの取得と利用のための体制

15.45-17.30: フィンランドと欧州連合: 北アフリカと中近東の地中海諸国における統計システムの発展を支援する技術協力プロジェクト連携諸国における体制

(スライド3) 第2日目

09.00-10.30: 共通の問題と解決策

10.45-12.30: 共通の問題と解決策

14.00-15.30: 参加国からの講演

15.45-17.30: 参加国からの講演

(スライド4) 第3日目

09.00-10.45: 照合入門

11.00-12.30: 参加国からの講演

14.00-15.45: 統計登録簿の行政的なデータソース 16.00-17.30: 参加国からの講演

(スライド5) 第4日目

09.00-10.30: 事例研究—フィンランドにおける行政的なデータソースの利用

10.45-12.30: 事例研究—フィンランドにおける行政的なデータソースの利用

14.00-15.00: 行政的なデータソースの分野における国際的な研究

15.00-15.30: 質疑応答

15.45-17.00: 閉会、感想と評価

(スライド6) 概説・課程の構成

- ・ 講演と事例研究—課程指導者と参加者
- ・ グループ演習と討論—オレンジ色スクリーンについての「警戒」
- ・ 質問時間—しかし不明な点があれば、尋ねてください。

(スライド7) 課程 (その1) 統計目的のための行政的なデータソースの利用

講師: スティーヴン ヴェイル (国際連合欧州経済委員会)

(スライド8) 行政的なデータソースとは何か

(以下、スライド8~スライド50は、本書の第1章、第2章および第5章のレジメであるから省略。)

本書を訳出する理由は、それが、実務者向けの概説的なテキストであって、スライドから伺われるように、指導的統計家による登録簿統計活動の国際的な「中間総括」と位置づけることができるからである。

わが国における登録簿統計 (register statistics) の研究は、1980年代に開始されたが、多くは紹介研

究やモデル的な研究に止まっており、行政登録簿から統計登録簿、さらには統計集団にいたる複合的な登録簿データの流れと変換過程に迫る本格的な研究は見受けられないのが現状である。その原因の一つは、これまで登録簿統計にかんする実践と経験がそれほど蓄積されておらず、統計調査と違って、統計登録簿が、政府省庁の行政登録簿と直接的に深く絡み合っているために、統計システムだけでなく、複雑な行政システムにかんする洞察が必要であるからである。本書は、章別編成に現れているように、登録簿統計が生成する主要な局面について具体的な考察がなされているから、専門的な調査研究者にとっても、示唆に富む好個の素材となっている。

概念や用語が熟成しつつある新しい分野であるだけに、訳語の選定には相当な困難があっている。例えば、基本用語である administrative source については、1) 政府の行政登録簿情報だけでなく、民間部門の「登録簿」データも包摂していること（本書第1章）、2) 編集者ヴェイル氏への照会によると、administrative data + meta data であることから、行政的なデータソースという訳語を当てている。また、register statistics, administrative register および statistical register については、日本において、register が各種の行政的な登録情報に関連して、登録簿と公称的に使用され、また法律学では登録簿という訳語が定着していることから、それぞれ登録簿統計、行政登録簿および統計登録簿と訳出している（注 ii）。その他の訳語についても、読者の検討をお願いしたい。

最後に、序、第1章～第3章、第5章、第7章～第9章を浜砂が、第4章と第6章を西村が分担して翻訳に当たっている。

訳者一同

(注 ii) 比較的新しい登録制度における登録簿の名称例：環境法規制登録簿（環境省）、簡易専用水道検査機関登録簿（厚生労働省）、（環境）国別登録簿（環境省、経済産業省）、マンション管理業者登録簿（国土交通省）、貸金業登録簿（財務省財務局各都道府県）、株券喪失登録簿（会社法規定）、宅地建物取引主任者資格登録簿（都道府県）、開発登録簿（厚木市）など。法律学における文献例としては、戸田典子「ドイツの連邦住民登録大綱法」『外国の立法』224号、2005年、および渡辺富子「【ドイツ】連邦住民登録法の制定」『立法情報』2013年7月号参照。

政府統計のための行政的なデータソースと第2次的なデータソースの利用
—原則と実務にかんするハンドブック—

スティーヴン ヴェイル (Steven Vale : 国際連合欧州経済委員会) 編著

注釈

本書でもちいられている名称と資料の表示は、いかなる国、領土、都市または地方の、さらにはその
の当局の法律的な立場にかんする、およびその国境や境界の画定にかんする国際連合事務局のいかな
る意見の表明も含んでいない。

謝辞

国際連合欧州経済委員会は、行政的なデータソースを統計目的のために利用するための国際的な研
修課程に大きく貢献した200名以上の参加者と、とくに以下に名前を挙げる客員講師に、感謝したい。
講師全員がこのハンドブックを、内容豊かにするために、たいへんに役立った。

ドリス アフザ (Driss Afza)、欧州連合：北アフリカと中近東の地中海諸国における統計システム
の発展を支援する技術協力プロジェクト

スー ファンデル (Sue Fendall)、イギリス国家統計局

リッタ ハララ (Riitta Harala)、フィンランド統計局

ジョン C. ヒューズ (John C. Hughes)、イギリス国家統計局

ベン ハンバーストーン (Ben Humberstone)、イギリス国家統計局

ペッカ マイルスカイライ (Pekka Myrskylae)、フィンランド統計局

カイヤ ルーオチャライネン (Kajja Ruotsalaine)、フィンランド統計局

版權の注意

このハンドブックの資料は、非営利目的のために、出所を下記のように明示することによって、複
写し、配布することができる。

出所：政府統計のための行政的なデータソースと2次的なデータソースの利用—原則と実務にかんす
るハンドブック、国際連合欧州経済委員会

ECE/CES/13

訳者注

(1) 編著者スティーヴン ヴェイル氏の略歴：イギリス国家統計局、欧州統計局や欧州協力開発機構
などにおいて、20年以上政府統計活動に従事した後に、最近7年は、国際連合欧州経済委員会統計
局において活動、現在、統計管理および近代化部の部長、統計作成と統計業務の近代化にかんする
高位の学識経験者グループの事務局長。その間、彼の主な研究領域は、企業調査、企業統計登録簿、
統計研修、行政データソース、企業家統計、メタデータ、普及と統計方法論。

序文

世界中の統計機関は、統計の作成過程の効率性を高め、とくに経費と人的資源を節約するように、いよいよ強まる圧力をかけられている。同時に、統計調査における回答者の負担を軽減する政治的な要請が大きくなっている。回答者が企業であるときには、とくにそうであって、それは、多くの政府が、官僚機構の縮小を、経済的な発展を支援し促進するための基本的な措置と見なしているからである。

このような圧力があるところでは、統計家は、データを収集する一つの方法である伝統的な調査方法に代わる方法を考案することを、いよいよ迫られている。おそらく、最も明瞭な解答は、利用できるデータが、既に、どこか他の場所に存在するかどうかを調べることである。多くの非統計的な機関は、いろいろな様式のデータを収集し、これらのデータは、統計調査によって収集されたデータの代替物に、そのままなることは、稀であるが、それは、しばしば、ときには複数のデータソースの組み合わせによって、完全に、または部分的に直接に統計データの収集に置き換わる可能性を開いている。

統計の作成過程において、行政的なデータソースが利用されている程度は、十分に機能する登録簿にもとづく統計システムを構築した国から、ちょうど、この方法を考慮し始めている国まで、国によって 差異は相当に大きい。

いくつかの課題については、特殊なテキストが存在しているが、これまでのところ、行政的なデータを利用する初期の段階にある国を支援する一般的な、国際的な方法論的な指針は存在していない。この手引き書の目的は、そのギャップを埋めることである。それは、統計目的のために行政的なデータソースを利用することにかんする国際研修課程のなかで、10年以上発展させられてきた資料にもとづいている。本課程は、ヨーロッパ全体、西・中央アジアと北アフリカの政府統計家である聴講者にたいして、10回以上提供されてきている。

課程は開催されるごとに、参加者と経験を共有し、参加者の応答を受け入れることによって、改善され、内容を高められてきている。それは、またフィンランド統計局やイギリス国家統計局から招いた、いろいろな専門家を客員講師として投入することによって、大きな恩恵を受けている。

スティーヴン ヴェイル (Steven Vale)、国際連合欧州経済委員会、課程指導者

訳者注：ヴェイル氏から、本訳書の刊行にたいして、序文の末尾に、つぎのような補追文が寄せられている。

「本ハンドブックの日本語版は、九州大学およびその雑誌である『経済学研究』の協力によって刊行されている。日本語圏の読者に本書の情報が利用できるように支援する寛大な協力を感謝したい。」

内 容 目 次

序文

内容目次

注釈

第1章 行政的なデータソースと第2次的なデータソースとは何か？

第2章 行政的なデータソースを利用する利点

第3章 行政的なデータソースを利用するための体制

第4章 一般的な問題と解決策（以上、本号）

第5章 精度と行政的なデータ

第6章 データの連結と照合

第7章 統計登録簿における行政的なデータの利用

第8章 統計調査を補完するための行政的なデータの利用

第9章 登録簿にもとづく統計システムに向けて

注釈

1) 参照文献にかんする注釈

このハンドブックは、他の論文、web サイトと出版書から、多くの引用・参照を行っている。これらの参照文献を求めたい者のために、インターネットのアドレスが可能なかぎり、与えられている。それは、執筆時にすべて確認しているが、それが、読書時に、なお機能していることを保証しない。読者が切断されたリンクに会ったら、support.stat@unece.org に、ご連絡下さい。

2) 練習問題にかんする注釈

第6章と第7章の末尾にある演習問題は、このハンドブックの基礎となっている課程から取られている。それは、これらの章で述べられている理論を補強するために、実務的な例を含んでいる。

第1章 行政的なデータソースと第2次的なデータソースとは何か？

1.1 はじめに

行政的なデータソースと第2次的なデータソースのデータを利用する実効性を考察する前に、その用語が意味することを明確に定義するために、いくらかの時間を割くことは、無駄なことではない。いくつかの定義が、現在利用できる文献に存在するが、そのなかの最も適切な定義を本章では検証する。本章は、比較的簡単に、適用範囲が広い定義を提案することによって、結びとなっており、それは本書おける残りの部分の基礎としてもちられている。

1.2 伝統的な定義

行政的なデータソースは伝統的には、統計以外の他の政府省庁が保持し、租税、給付やサービスについて執行するために、収集・利用されているデータの収集と定義されてきた。おそらく、最も包括的な伝統的な定義は、カナダ統計局の Gordon Brackstone によって、彼の1987年論文「行政データの統計的問題：問題と挑戦」において、設定された¹⁾。Brackstone は、行政的なデータを区分するために、つぎのように4つの特徴を特定した。

1. 統計機関にデータを提供する機関とデータにかかわる機関は、(多くの統計調査とは対照的に)異なっている。
2. データは、もともと一定の非統計的な目的のために収集され、その目的がデータソースの部局における処理に影響する。
3. 目的母集団を完全に捕捉することが、その課題である。
4. 行政的なデータが収集され、処理される方法の管理は、行政機関の裁量である。

この定義は、統計データとメタデータ交換プロジェクト (SDMX) によって提案された方向に、おおむね沿っている²⁾。

「単一または複数の行政的な法規を施行する目的のために、収集され、保持されている情報を含むデータ保持体 (*data holding*)。』

1996～97年の期間に、欧州統計局内部の専門調査団 (task force) が、いろいろな統計分野にまたがる行政的なデータソースの利用にかかわる作業を、よりよく調整する方法を検証した。この専門調査団は、行政的なデータソースの定義方法を考察するために、データソースを簡単に類別する方法をもちいた。最初に、すべてのデータソースは、第1次的なデータソース (統計目的のために収集されたデータ) と第2次的なデータソース (他のすべてのデータ) に区分された。広義の定義が、民間部門のデータソースを包括しているのにたいして、行政データベースの伝統的な、または狭義の定義は、

1) Brackstone GJ : "Statistical Issues of Administrative Data: Issues and Challenges", in "Statistical Uses of Administrative Data - An International Symposium", organised by Statistics Canada, 23-25 November 1987 (Proceedings published by Statistics Canada, Ottawa, December 1988)

2) www.sdmx.org 参照

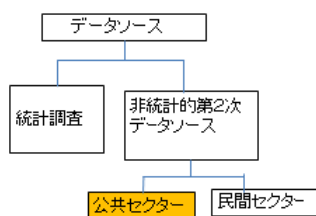


図1.1 狭義の定義

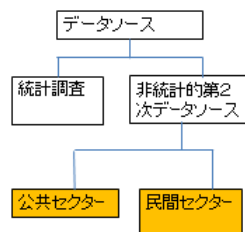


図1.2 広義の定義

公共部門の非統計的なデータソースだけを含んでいる。

広義の定義方法は、欧州統計家会議が、公刊書「統計的なメタデータにかんする述語」において採択した行政的なデータの定義と整合的である³⁾。

「統計機関以外のデータソースによって収集されたデータ。」

狭義と広義の定義は、図1.1と図1.2のように表示することができる。

したがって、狭義の定義では、行政的なデータソースは、第2次的なデータソースの部分集合であるが、広義の定義では、この用語は、同義である。

広義の定義を支持する理由が多くなっている。それはつぎのような理由を含む：

●政府活動の民営化の拡大：

いくつかの国では、以前には政府省庁が実施していた法規の施行が、民間または半民間の組織に移されてきている。典型的な例は、通常、健康、教育または公共事業サービスであって、以前は国家が独占していた事業が、民間企業や非営利団体の事業に、ますます置き換えられている。

政府部門の支援による行政登録簿の運営を含む登録業務が、いくつかの国では民営化に向けて検討中である。それは、公共部門と民間部門の機能にかんする伝統的な区分がいよいよ鮮明でなくなり、行政的なデータソースにかんする伝統的な、または「狭義」の定義が限定的でありすぎる定義になって行っていることを意味する。

●民間部門のデータ、および「付加価値を付け加えて再販売する業者」が増加していること

電子化された情報の量が世界的に指数関数的に増加しており、ほぼ5年ごとに10倍の比率で増加している。この「データの洪水」の小さな部分だけでも、政府統計にとって重要であるならば、それが包含するデータの量と事項の範囲は、なお膨大である。

同時に、データの商業的な価値が明確になりつつあり、データの市場が民間部門のなかで急速に拡大している。それは、市場調査目的の住所リストの開発と販売に始まり、信用格付けデータと経済情報の提供を捉えるまでに発達して、現在、実質的にあらゆる種類のデータを包含するまでに広がっている。この市場の大きさが増大すると、それから利益を求める企業数も増大してきた。民間部門は、

3) www1.unece.org/stat/platform/download/attachments/9110092/Metadata+terminology+2000.pdf?version=1 参照。

データが、たいへんな価値をもつ商品であることを実証している。

比較的近年の発展は、データ市場において民間部門「(データに)付加価値をつけて再販売する事業」が出現したことである。この事業では、いろいろな公共部門や民間部門のデータソースから、既存のデータを得て、それを組み合わせて、整理し、ときにはその有効性を保証して、他の組織に再販売する。例えば、Dun & Bradstreet, Bureau van Dijk & Hoppenstedt Bonnier のような経済データを販売する事業者が含まれている。

このような種類のデータソースは、民間部門のデータ供給者が、即時的にデータを処理することができ、たんに、費用を多くの顧客の間に分散できるという理由から、そのデータを、統計機関よりも安く提供することができるときには、政府統計の供給者にとって重要である。企業集団にかんする欧州統計登録簿を開発する「ユーログループ」プロジェクトは、まさにそのような理由から、このようなデータソースを利用している。

そのようなソースのマイクロデータを直接的に利用することにたいする代替法は、基準指標を作成するために集計データを利用することであって、民間部門と政府統計登録簿の目的集団の捕捉範囲を比較する。イギリスにおける企業統計登録簿の捕捉範囲を、代表的な民間部門のデータソースのそれと比較する試行は、都心部や休日の行楽地における経済活動にかんする統計の過小把握を示し、政府統計における限界のかつ季節的な活動を捕捉することにもなう困難さを例証し、このような過小把握の大きさを明示する指標をも与えている⁴⁾。

●新しい種類のデータにたいする利用者の関心

政府統計の利用者は、新しい種類のデータを常に要求している。統計調査における回答者の経費と負担を減少させる圧力が、このような需要に対応する統計調査を、新規に開始することを困難にしているから、統計家は、いよいよ代替的な解決法を求めることが必要になっている。民間部門におけるデータソースの量、内容および捕捉範囲が大きくなっているから、統計調査の代替法としての魅力も増している。

1.3 行政的なデータソースの事例

前節で議論したように、統計目的のために利用できる行政的なデータソースの潜在的な範囲は、大きくかつ成長している。つぎのリストは、すべてを尽くすことを意図したのではなくて、行政的なデータソースの操作的な定義に到達するための最終的な段階として、潜在的なデータソースの範囲と種類を示そうとしている。

- ・ 租税データ
- － 個人所得税
- － 付加価値税 (VAT)

4) 試行の結果は、捕捉範囲の地図の様式で論文「イギリスにおける小地域企業統計の開発」に示されている。つぎのインターネットアドレスで利用できる：

<http://live.unece.org/fileadmin/DAM/stats/documents/ces/sem.53/wp.7.e.pdf>

- 営業税／収益税
- 財産税
- 輸入・輸出関税
- ・ 社会保障データ
- 社会保険料
- 保険給付
- 年金
- ・ 保健・教育記録
- ・ 個人・企業・財産・自動車の登録システム
- ・ 身分証・旅券・自動車免許証・選挙人登録簿
- ・ 農家登録簿
- ・ 地方議会登録簿
- ・ 建築許可証
- ・ 免許システム、例えばテレビ、特定財の販売
- ・ 公刊企業勘定
- ・ 企業が保持する内部会計データ
- ・ データを保持する民間企業
 - 信用機関
 - 企業分析者
 - 公益事業企業
 - 電話帳
 - 顧客カードをそなえる小売業者など

1.4 要約

結論として、本章は行政的なデータソースと第2次的なデータソースの広義の定義にかんする事例を議論している。それは、新しい様式のデータソースの潜在的な価値について、創造性に富む評価を行う必要性を強調する。そのために、行政的なデータソースと第2次的なデータソースの定義は、統計家にどんな人為的な制約もおくべきでなく、できるだけ広範であるべきである。ゆえに、用語「行政データソースと第2次的なデータソース」は同義語であるから、本書では、今後、二つの概念を包含するために用語「行政的なデータソース」を使用する。

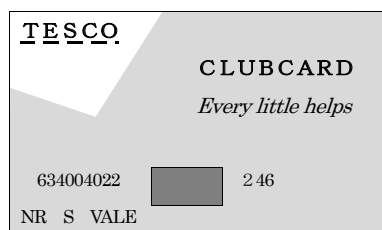
したがって、提案される定義はつぎのようである。

行政的なデータソースは、基本的には統計目的のために収集されない情報を含むデータの保持体 (*data holdings*) である。

この定義は、本書では、この後の部分における基礎としてもちいられる。

囲み資料1.1 将来に向かって：顧客カード（Store Cards）－潜在的なデータソースか？

顧客カードは、民間部門における新しい種類のデータソースの典型的な例である。割引や独自の特殊なサービスを提供する便益にたいして、顧客カードの利用者は、それを利用するごとに、販売業者に多くのデータを提供する。顧客カードを所持していると、販売業者は、あなたについて、つぎのようなデータを得る、または導き出すことができる。



- ・氏名、住所、性、年齢
- ・家庭環境（例えば、あなたが、定期的に乳児用品、おもちゃ、ペットフード、または一定の量や大きさの食肉のような製品を購入するとき、あなたの世帯構成を推測することは容易でしょう）
- ・仕事上の地位と所得にかんする指標（例えば、あなたが購入する時間は、あなたが仕事に就いているかどうかを示す、また購入する財の種類は、可処分所得の指標である）。
- ・自動車所有（ガソリンの購入、車手入れ用製品）、宗教（特定の宗教に結びつく用品の購入、例えば、イスラム教やユダヤ教の教義によって、食べることが認められている肉）などのようなその他の世帯指標など。

これは、潜在的なデータソースの極端な例とおもわれ、近い将来に政府統計の目的のためには考慮に入らない事例とおもわれるかも知れない。しかし、いくつかの国では、小売り販売・価格にかんするデータソースとして、大規模な小売業者のレジロールデータの利用が考慮されており、ニュージーランド統計局は、電子カードの取引データをもちいて実験的なデータ系列を作成してきている⁵⁾。

とくに顧客カードの捕捉範囲が、他の商業的なソースからのデータのみならず、異なる様式の顧客カードのデータと連結されることによって改善されるならば、その利用は、必然的につぎの段階と見なすことができよう。もし、このような種類の行政的なデータソースが政府の統計家によって無視されるならば、このようなデータを取得する民間部門の事業が、人口センサスデータのような基本的な政府の統計に代わる、妥当かつ、より費用効果的な代替財を提供し始めるまでに、どれくらいの時間がかかるだろうか。

5) http://www.stats.govt.nz/browse_for_stats/Corporate/Corporate/nzae-2007/~media/Statistics/Publications/NZAE/The%20development%20of%20electronic%20card%20transaction%20statistics/development-of-ect-statistics.ashx

第2章 行政的なデータソースを利用する利点

2.1 はじめに

前章では、行政的なデータソースの性格と範囲を定義したが、そのソースが統計家の関心を集める理由については、実質的に考察しなかった。本章は、行政的なデータソースによって、統計のデータソースを補完するか、それを置き換えるために、政府統計において、行政的なデータソースを利用する多くの潜在的な利点を考察する。もちろん、それは、すべて良好な情報ではなく、利点とともに、また一般的に、克服されねばならない広範な問題もある。このような問題と解決方法は、第4章において取り上げる。

2.2 経費

統計調査は、データを収集する高価な方法である。調査票が開発され、(特別な抽出フレームの創案さえ要求されるかもしれない) 標本が設計されなければならない、回答者が接触され、可能ならば、彼らの回答を勧めるために再接触がなされなければならない、回答は処理かつ確認され、そして調査結果が集計されなければならない。電子計算機が、データ処理の負担の多くを取りさるが、行程の全体が、なお労働集約的であって、とくに回答を求める過程はそうであって、おそらく完全に自動化することはできないであろう。

伝統的なセンサスは、きわめて大きな規模で実施されるから、より一層そうである。なお、国民、企業、農家などについて、伝統的なセンサスを実施している国家の統計機関は、しばしばそれを実施するためには、経常予算でまかなうには経費が大きすぎるから、特別な財源を要求する。それは、政治家たちには非常に眼に着きやすく、そのために政策的な優先度の変更を求められやすい弱点となっている。

統計を作成するために行政的なデータソースを利用する立ち上げ経費は、もちろん統計調査の立ち上げ経費と同様に高いが、運営する経費は、一般に相当に低い。つぎの表2.1と図2.1は、欧州連合諸国において、2000～2001年に人口センサスを実施した経費を示している。センサスが、全体として行政的なデータソースにもとづいているフィンランドと、伝統的な書式の調査票が使用されたイギリスやオーストリアのような他の国々の間にある人口1人当りの経費の大きな差は、行政的なデータソースの利用を拡大していくためには、最強の論拠である。

行政的なデータソースの利用は、しばしば料金が無料である。そのデータの出所が公共部門であるときには、とくに、そうである。例えば、公共部門からのデータの抽出や送信、または民間部門のソースからのデータの購入に料金が必要なときでさえ、しばしば、同一の情報を、調査によって収集するよりも、行政的なデータを使用する方が、なお安価である。

統計調査が、まだもちいられているところでは、効率的、かつ正確な標本抽出フレームが必要である。その標本抽出フレームを作成するために利用される統計登録簿は、しばしば大きくかつ複雑であるために、統計調査、またはセンサスのデータをもちいて、十分に、統計登録簿を適切に受け入れ、

表2.1 いくつかの欧州連合諸国における人口センサスの経費

国名	総経費 (単位：百万ユーロ)	人口1人当経費 (ユーロ)
ベルギー	24	2.3
ギリシャ	50	4.5
スペイン	167	4.1
フランス	248	4.1
アイルランド	44	11.2
イタリア	298	5.3
ルクセンブルク	5	10.6
オーストリア	56	6.9
ポルトガル	46	4.5
フィンランド	0.8	0.2
イギリス	367	6.2
ノルウェー	15	3.3
スイス	99	13.6
チェコ共和国	80	7.8
エストニア	10	7.4
ハンガリー	40	3.9

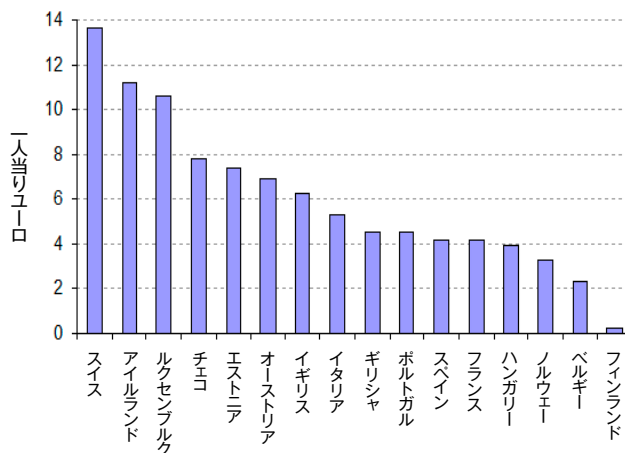


図2.1 人口1人当りの人口センサスの経費の比較

(出所) 表2.1と図2.1ともに欧州統計局出版：(「EU、EFTA と加盟候補国の2000ラウンド人口・住宅センサスにかんする資料」⁶⁾)

6) http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-CC-04-002/EN/KS-CC-04-002-EN.PDF

それを維持することは、たいへん困難かつ高価である。ゆえに、行政的なデータが統計調査に置き換わらないときでさえ、行政的なデータは、統計登録簿を受け入れ、維持するために使用されることができて、総経費を減少させることに役立つことができる。

2.3 回答の負担

行政的なデータソースは、データ提供者の回答負担を軽減させることに役立つ。多くの国において、それは、非常に政策的に考慮されており、回答者が企業であるときには、とくにそうである。経済の発展と成長を促進する政策は、しばしば規制的な負担を減少させることを含んでいる。このような環境では、統計調査は、しばしば削減が容易な対象と見なされる。

企業側は、一般に、そうすることを好まなくても、登録目的と課税目的のために、データを提供する理由を理解する。しかし、彼らは、しばしば統計データの要求を余計な、あまり必要ではない負担と見なす。彼らが、すでに他の政府省庁に情報を提供しているときには、国家の統計機関から、類似の要求を受けることに不快感を示すようになるかも知れない。したがって、政策者と回答者が、統計にたいする回答負担の軽減を求めるために結束するときには、国家の統計機関にとって、その圧力に抵抗することはきわめて困難であって、他の機関が収集したデータを再利用することが、必然的な解決法となる。

2.4 頻度

経費と回答負担の軽減に関連して、行政的なデータソースを利用するもう一つの利点は、いくつかの事例では、それが、追加的な回答の負担をとまわず、ほとんど経費の追加がなくても、もっと頻度が高い統計作成を許容していることである。それは、フィンランドの事例であって、伝統的な方法もちいる国が、5年ごとにか10年ごとにしか、人口センサスのデータを作成することができないのにたいして、フィンランドでは、行政的なデータソースからセンサスデータが、年次ベースにおいて作成されている。

行政的なデータによって作成される統計の頻度にたいする主な制約は、通常行政的なデータソースが更新される頻度である。したがって、行政的なデータが、季節的な偏りが無い時間的なベース(rolling basis)で更新されないかぎり(あるいは少なくとも季節的な偏りを分離するための十分な情報がないかぎり)、年に1度更新される行政的なデータから、月次の統計を作成することは困難であろう。

しかし、事象(例えば、出生、死亡、計画許可の認証)を記録する(record events)データソースがそうであるように、特定の期間に基礎をおかない行政的なデータソースは、しばしば相当に弾力的である。それは、事象の日付が正確に記録されているかぎり、行政的なデータソースが、与えられるいかなる期間についても、すなわち日次に至るまでの周期についても、統計の作成を許容するからである。

2.5 捕捉範囲

標本調査が直接には比較的（母集団の）小さな部分を、直接に捕捉することだけができるのについて、行政的なデータソースは、完全に、またはほとんど完全にその目的集団を捕捉する。したがって、行政的なデータソースの利用は、調査誤差をなくし、非回答を除去（ないしは相当に縮小）し、いろいろな部分集団について、たとえば地理的に小地域における、またはその他の特定の特性をそなえる回答者層について、もっと正確かつ詳細な推定値を提供する。

2.6 適時性

行政的なデータソースの利用は、一定の変数にかんして、より最新の情報を取得することを許容することによって、統計の適時性を高めるかも知れない。それは、統計調査が、一般的には、母集団を分析し、かつ標本を最適化するために、企画し、調査票を設計し、それをもって試験調査を行うために時間を要するからである。それは、とくに年次的なまたは特定のデータ収集に適はまる。したがって、適切な行政的なデータソースを利用することが、より有効な解決法である。しかし、行政的なデータソースの利用が、とくに短期指標については、適時性を低下させる場合もあるかも知れないことを指摘しておくべきである。

行政的なデータソースが適時性に、とくに積極的な影響を与えることができる一つの分野は、統計登録簿と調査フレームの管理である。明らかに上述した捕捉範囲にかんする利点から、目的母集団の変動（例えば、人や企業の出生と死亡）にかんする行政情報は、しばしば調査情報よりも、非常に更新度が高い。

2.7 国民の心象

データの共有、とくに異なる政府省庁間のそれにかんする世論は、国によって相当に異なっている。世論がデータ共有を受容する、ないしはそれに好意的であるところでは、既存のデータソースの利用を増加させることが、データ共有をより効率的かつ費用効果的にすることによって、国家の統計制度の威信を高める。

しばしば、国民の間には、データ共有にたいして、一般的な不安感が存在しているが、また政府の効率性を改善する要求のような矛盾する圧力もあって、とくに、それが選挙民に人気がある減税、または保健や教育のような分野における予算の増額をもたらすときには、とくにそうである。「政府への参加」と言うような政治的な標語は、しばしば国民に訴え、プライバシーの喪失にたいする恐れに対応する助けとなる。したがって、行政的なデータソースの利用にとって、国民の心象を向上させることが利点と見なされる程度は、その利用が、どのように国民に呈示され、それが国民に受け入れられるかに大きく依存する。

第3章 行政的なデータソースを利用するための体制

3.1 はじめに

行政的なデータソースからデータを取得することが、統計目的のためにそのようなデータの利用を拡大することにたいする基本的な障壁の一つになっている。本章は、いくつかの国の事例と経験を参考にすることによって、行政的なデータソースの利用を容易にするために必要ないろいろな体制について述べる。これらの体制には、一般的に、いくつかの次元がある。法律的、政策的、組織的、および技術的な体制がそれであって、以下において考察される。行政的なデータを利用する利益が実現される前に、これらのすべての領域で、合意に達することが必要である。

3.2 法律的な体制

法律的な体制は、通常、国家的な次元において構築され、とくに国家的なデータソースと環境によって特徴付けられている。しかし、いくつかの事例では、また副次的な国家（例えば、州）の次元、または国際的な次元において関連する法律があり得るだろう。後者の一つの例は、構成国を拘束する欧州連合の統計にかんする規約である。そのような場合には、行政的なデータにたいして、二つ以上の代替的な法律的な方途があり得る。

多くの国家の統計機関は、その役割と責務を規定する法文書を、典型的には統計法の形態において、そなえている。多くの国では、この法文書は、行政的なデータを取得する特別な条項を含む。つぎの例は、アイルランド⁷⁾とノールウェー⁸⁾の統計法である。

囲み資料3.1 アイルランドの1993年統計法からの抜粋

第30条項 (1) 本法のもとで、活動する（統計）局を支援するために、統計局長は、いかなる公共機関にたいしても、通知書の送付によって、つぎのことを要求することができ、公共機関は、本条項の小項 (2) にしたがって、そのようないかなる要求にたいしても、無料で応じなければならない。

- (a) 統計局員に、適正な時間ならいつでも、当該機関の管理下にあるいかなる記録も閲覧し、審査し、記録を複写し、または抜粋することを認めること、
 - (b) 統計局のいかなる職員でも、そのような要請をするときには、いかなる記録の複写または抜粋も、統計局に提供すること。
-

第31条項 (1) 統計局長は、いかなる公共機関にたいしても、その機関が保持する記録が統計情報のソースである可能性を評価するために、および統計目的のために記録方法と記録システム

7) www.irishstatutebook.ie/1993/en/act/pub/0021/index.html (See sections 30 and 31)

8) www.ssb.no/english/about_ssb/statlaw/statlov_en.html (See chapter 3 - 2)

を開発することが適切かつ、実効性があるときには、それを統計局長と協議し、彼に協力することを求めることができる。当該の公共機関は、財源が許すかぎり、そのような要求に従わなければならない。

(2) いかなる公共機関も、情報を保管し、訂正するためのいかなるシステムについても、それを導入し、修正し、または拡張するために、または統計調査を実施するために、提案を行うときには、統計局長と協議し、後者が、提案に関連して、合理的に行うどのような勧告をも受け入れなければならない。

いくつかの国家の法律的な体制は、統計目的のために行う行政的なデータの取得について、他の国よりも、強い権限を与えている。それは、国家的、歴史的、政治的かつ文化的な要因が法律的な体制に強く影響しているからである。文化的な要因はとくに重要であって、いくつかの文化は、他の文化よりも、政府省庁間においてデータを共有する理念について非常に肯定的である。このように国々に相違が存在するために、法律的な体制は、とくに調整されないか、整合的でない。

この整合性の問題に取り組むために、欧州連合は、一般に「統計法」⁹⁾として知られている欧州統計にかんする規約第223/2009号に、行政的なデータの取得にかんする規定を含めている。この規約は、欧州の統計法規のもとで、構成国の統計機関に、この責務に対応するために必要な行政的なデータを取得する権利を与えているが、そのような取得は、なお国家の制約と条件に従うと述べている。

特殊な統計の領域における欧州連合の個別的な規約は、さらに進んでおり、国家的な限界や条件にたいする依存性を離脱している。この一つの例が、企業登録簿規約であって、規約の要求に応ずるために、行政的なソースのデータが必要であるときには、いかなる行政的なデータソースにたいしても、制限がない利用を定めている¹⁰⁾。

行政的なソースからのデータを利用することを規定するとともに、法律的な体制はまた、そのような利用、すなわち行政的なデータの利用に制限を設定する。しばしば、データは、特定の統計目的のためにだけ利用でき、個々の記録の機密性が保持されなければならないという制約が設けられている。

例えば、法人化されていない企業、とくに企業データが企業の所有者に関係する個人データと見なされるときには、単一所有者の企業にかんするデータの利用には特別な制約が設けられている。ここでは、企業の利益は、個人の所得に等しいと見なすことができる。個々の市民にかんする情報を包含するデータ保護法がある多くの国では、そのような事例について、企業データと個人データを形作る区分を明確にすることが重要である。

9) Article 24 of Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics : <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:087:0164:0173:EN:PDF>

10) Article 4 of Regulation (EC) No 177/2008 of the European Parliament and of the Council of 20 February 2008 establishing a common framework for business registers for statistical purposes : <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:061:0006:0016:EN:PDF>

立法過程は時間を要し、統計は、しばしば、立法者によって優先度を比較的に低く見られているから、行政的なデータを使用する利益を訴え、強調し続ける期間が必要になるかも知れない。通常、統計法規を導入し、または改正するために、あらゆる努力がなされて、多くの機会を作ることが必要である。とくに現在の要求にかなう法律を提案する過ちを避けることが重要である。法律を改正するつぎの機会が来るまでに10年以上を要するかも知れないから、行政的なデータを利用するためには、長期的な戦略を立てることと、立法の提案が、予見できる将来を見通したすべての要求に対応することが確保される必要がある。このようにして、立法は、短期的には障害と見なすことができるが、長期的には好機である。

立法が障害として止まるとしても、それは、必ずしも、行政的なデータのいかなる利用をも抑制しない。一つの手本となる事例では、法人税データを取得する適切な法律的な体制作りを待っている間に、イギリスの国家統計局の1人の職員が、統計目的のために、このデータを利用する実効性を探るために、臨時的に税務庁に転勤した。この職員は、税務庁に勤務する間にマイクロデータを利用して、物理的には、税務庁の庁舎内において作業し、統計局に、機密が露見しない集計された分析結果だけを持ち帰った。この方法では、取得を実現する法律的手続きを探りながら、税務データの実質的な価値を適切に評価する課題を含めて、いろいろなデータ問題に着手することができた。

法律的な制約は、しばしばマイクロデータの利用、すなわち個人や企業にかんする情報の利用に関連することに留意すべきである。統計家は、習慣的には集計結果を作成するために、この次元のデータを処理してきたが、ときには、代わりに機密が露見しない、低い集計レベルのデータを取り扱うことが、実効性があるかも知れない。いくつかの事例では、それは、統計単位を、個人的な次元から一定の特性を共有する小さな個人集団の次元に移して単に再定義すること、おそらくは、その集団の構成員数に等しい加重をほどこすことによってなされた。

3.3 政策的な体制

多くの国では、政府内部のデータ共有にかんする一般的な政策がとられ、それが統計目的のために行政的なデータを取得する権利に影響を与えている。しかし、法律を改定するよりも、政策を変更することが、しばしば容易であって、政策は、時とともに進展する傾向がある。したがって、国家の統計機関は、完全に政策形成に関与し、政策の変更をもたらす政府内のいかなる審議にも積極的に参加することが重要である。このようにして、統計システムに最大限、可能な利益が与えられるように、いかなる変更もなされるべきである。

また、政策的な体制は自発的な活動規程 (codes of practice) を含む。そのなかで、統計目的のために、最も重要な規程は、国際連合の「政府統計の基本原則」である¹¹⁾。原則5は、経費効率性に関連し、行政的なソースからのデータ利用について、つぎのように提案する。

「統計目的のためのデータは、統計調査であれ、行政記録であれ、あらゆる種類のデータソースから

11) <http://unstats.un.org/unsd/methods/statorg/FP-English.htm>

獲得することができる。統計機関は、精度、適時性、経費と回答者の負担を考慮することによって、データソースを選択するべきである。」

また、原則5にかんする注釈は経費効率性を強調し、さらに、つぎのように述べている。

「統計局は、経費節約的でなければならず、精度、適時性、経費と回答者の報告方法を調整することによって概念、データソースと方法について、最良の選択を行う…。統計局の全体的な経費効率性は、組織的な計画と操作、統計方法の堅実な応用、情報・通信技術の完全な利用、および行政記録の取得によって影響を受ける。」

欧州統計活動規程¹²⁾には、同様な条項があるが、行政的なソースからのデータ利用は、少し異なった文脈において奨励される。原則2は、データ収集の義務に関連し、つぎのように述べている。

「統計当局は、欧州統計の目的のために、情報を収集することについて、明確な法律的委任を受けていなければならない。行政、企業と世帯、および国民全般は、統計当局の要求に応じて、欧州統計の目的のために、データの取得ないしは提供を許容する法律に、強制的に従わされる。」

原則9は、統計調査にたいする回答者の負担が過大ではないことを保証することにかかわっている。それはつぎのように述べる。

「報告の負担は利用者の（データ… 訳者）需要に釣り合っているべきであって、回答者にとって過大であってはならない。統計当局は、回答の負担を監視し、それを軽減する時間的な目標を設定する。」この原則が適用される程度を測るために提案された指標の一つは、つぎのものである。

「行政的なデータソースは、できるかぎり、情報にたいする要求の重複を避けるために利用される。」

活動規程は、国家次元でも存在しているが、それは、データが特定の合理的な目的のためにだけ利用されることを、国民に確信させる方法として、しばしば有意義である。なんならかの実質的な有効性を得るためには、典型的には国家の統計機関におけるインターネットサイトによって、活動規程が一般国民に活用できるようになっていることが、重要である。

3.4 組織的な体制

ひと度、法律的な体制と政策的な体制が、行政的なデータの利用を許容するようになると、データの流れを容易にする組織的な調整を考慮することが必要になる。一般的には、それは、文書による協約の様式を取る。とくに民間部門が含まれるときには、それは契約であるかも知れないが、協定が政府省庁間で結ばれるときには、それは「事務次元の合意」、「議定書」や「協約」であるだろう。相違は、協定には法律的な拘束力がないのにたいして、契約には、拘束力がある傾向がある。

そのような協約には、一定の基本的な特徴がある。それは、つぎのような特徴である。

- ・ 法律的な根拠：統計目的のために行政的なデータソースを利用することを許可する法律と、この利用を制限する法律からの引証
- ・ データの譲渡者と受取者の氏名：行政機関と統計機関双方において、データ譲渡にかかわる責任者

12) http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/code_of_practice.

の氏名、交渉内容と職名が記録されるべきである。それはいくつかの事例では、データを利用しないしは閲覧する資格がある統計機関内のすべての者を含むことができている。

- ・含まれるデータの詳細な記述：それは、含まれるデータセットと、それが含む変数名を確認する情報を含む。
- ・データ提供の頻度：それは行政機関がデータを供給する時点と頻度を特定する。
- ・精度の基準：これは供給されるデータの精度にかんするパラメーターを設定する。受け取るデータが目的に適合していることを保証するために、例示が、一定の基準に適合する対応措置にたいする要求、または欠落する変数または誤った変数の割合の最高限度にたいする要求を含むことがある。異なる変数に割り当てられる優先度、したがってまた精度の確保に投入される尽力度は、行政機関や統計機関によって異なり、共通の基準について協約することは困難である。
- ・機密性の規則：データの利用目的、開示を制限する規則と手続き、およびデータが統計機関の利用者に譲渡される環境を設定することが重要である。
- ・技術的な基準：その詳細は、次節のような技術的な体制にかんする条項に含まれる。
- ・メタデータの規定：供給されるデータに、日付け、使用されているすべての符号の記述、使用される単位にかんする情報など、関連するメタデータが添えられていることが重要である。
- ・データ供給にたいする代価の条項：政府の省庁間におけるデータの譲渡は、一般的には無料であるが、いくつかの案件では、統計機関は、データの抽出や転送にたいして分担金を求められるかも知れない。民間部門における機関のデータは、とくに、そのデータソースの利用者が政府部内に複数あるときには、割引を交渉することができるかも知れないが、しばしば市場相場で課金される。いくつかの事例では、受け取るデータの代価として、統計の分析結果や専門的な知識を提供することがある。
- ・協約の期間：協約は一定の期間であるかも知れないが、もしそうであって、必要ならば、それを更新、または延長する条項を含むべきである。もう一つの方法は、関係機関の一つが変更を望む期間まで有効であると見なす協約を結ぶことである。
- ・環境の変化による不測の事態：行政的なデータソースに影響する変化にかんする事前の予告を受けることが、統計機関にとっては、重要である。協約には、統計にたいする変化の影響を最小限にするために、予定されている変更は、どのようなことでも、できるかぎり早く、統計機関に通知されることを明記すべきである。
- ・争議を解決する手続き：協約は、通常、上級管理者、または場合によって関連大臣のところまでに、問題が拡大する可能性がある統計機関と行政機関の間の係争を解決する方法を明記すべきである。

3.5 技術的な体制

技術的な体制は、データが譲渡される技術だけでなく、関連するすべてのデータやメタデータの基準に関係する。データ譲渡の技術は、郵便による文書記録の送付から安全な電子的な伝送による実時間的な更新に至るまで、いかなる様式も取ることができる。利用される技術については、データの

譲渡機関にもデータ受取機関にも支障を来さない技術的な可能性が考慮されなければならない、しばしば、少なくとも双方のうちの一つの機関にとって、次善の解決法を考慮する妥協が行われる。

データやメタデータを送信するために、数多くの国際標準版が存在する。名称を少しだけ挙げると、XML、SDMX や DDI などである。また、いくつかの国では、とくに政府部内におけるデータ譲渡のために、国家版がある。したがって、使用する標準版を定めておくことが重要である。

3.6 要約

統計目的のために行政的なデータを利用することを許容する法律的な体制があることが、不可欠である。上述した他の体制は不可欠ではないが、データの円滑な流れを保証し、データ供給者と統計機関の間に発生する問題や誤解を最小にするために、たいへん有効である。この理由から、それが、すべての機関によって合意された書式の文書に配慮されているなら、有用である。

諸国間の活動を比較することは、基準となる目標を設定するために有用であるが、特定の国家的な事情や問題が特有の解決法を要求することを想起すべきである。国際的な標準版は、指針を提供する観点において有用であって（それを遵守することは政策的な目標と考えられる）、それは可能なかぎり、行政省庁との審議にかけられるべきである。

囲み資料3.2—事例研究：イギリスの体制

・ 法律的な体制

2007年統計・登録業務法は、行政的なデータを取得する体制を整えたが、他の多くの国におけるようなデータ取得にかんする包括的な権限を与えなかった。行政的なデータを取得し、利用する条件は、1994年付加価値税法のように、しばしばデータソースが特定された立法によって決定されている。新しい行政的なデータソースを利用することは、議会の承認にしたがう。イギリスは、また欧州連合の構成国として、行政的なデータソースの利用に関連する欧州法の条項にもしたがっている。

・ 政策的な体制

国際連合の政府統計における基本原則と欧州統計システムの活動規程の適用に加えて、政府の統計業務を行う機関にたいする国家活動規程¹³⁾がある。統計目的のための行政的なデータの利用に関連する基本的な条項は、つぎのようである。

・「5 (f) 行政的なデータソースから得られたデータには、統計目的のために特別に収集されたデータと同じ機密性の基準が適用される。」

・「7 (c) 国の統計作成における行政的なデータソースの意義が認知され、統計目的が、行政システムの設計において活かされるべきである。」

13) http://www.statistics.gov.uk/about/national_statistics/cop/default.asp

・「7(d) 統計システムは、データ統合によって付加される価値が最大になるように設計される。」
活動規程は、いろいろな議定書によって支援されており、それは、つぎのような記述を含む回答負担の管理にかんする議定書¹⁴⁾を含んでいる。

・「2 新規の統計調査は、既存のデータソースと重複しない…。国の統計作成者は新規の調査を導入する前に、既存の調査データ、行政データ、およびその他の調査以外のデータソースを考慮する…。調査は、適切な他のデータソースが存在しないときにだけ実施される。」

・「4. 国の統計作成における行政的なデータソースの意義が認知され、統計目的が、行政システムの設計において活かされるべきある。国の統計は、適切ならば、政府企業や公益事業体を管理するために供給される情報から得られる。それは、できるかぎり、行政を支援するシステムから関連データを直接に抽出することによって、達成される。国の統計作成者は、行政システムが経済的な方法において、統計目的のためにデータを得ることができるように、行政システムの設計を所轄する組織に影響を与えることを求める。」

・組織的な体制

政府省庁間においてデータを譲渡するための組織的な体制は、「業務次元の協約」に具体化される傾向がある。それは、上級管理者の次元において調印されるが、法律的な拘束力はない。それは、協約の中心部分に一般的な条項を含み、補則において、特定のデータにかんする要求と細目を詳説する。通常、代価の支払はないが、いくつかの事例では、統計の分析結果または統計方法が、返礼として提供される。

企業統計登録簿のために、行政的なデータを提供する政府省庁は、登録簿運営委員会に代表を送っており、同委員会は、利用者も含んでいる。それは、データの利用方法とデータ精度の意義にかんして、理解を高めることに役立っている。

企業登録庁（企業省）は、より商業的なベースにおいて業務を行っており、登録庁からのデータの譲渡は、契約によって、代価を支払う体制を取っている。企業の所有と経営体にかんするデータもまた、民間部門における経済データの供給者から購入している。

・技術的な体制

多くのデータ譲渡は、固定長のフィールド、または標準的な区画のテキストファイルによって、行われている。一定の標準化ができていない地域では、企業住所の様式で行われている。それは、郵便局の標準方式にもとづく住所参照ソフトウェアを利用することによって促進されている。

多くのデータ譲渡が今日、郵送のフィロピイデスクによって、少量のデータセットについては、安全な電子メールネットによって行われている。しかし、企業統計登録簿において利用されている付加価値税データについては、毎日、更新されるシステムが、政府部門内の保護ネットワーク

14) http://www.statistics.gov.uk/about/national_statistics/cop/downloads/respondentload.pdf

によって送られる交信ファイルをもちいて設定されている。

メタデータは、少ない頻度で、データに添付するか、データと分離して参照表として送信されている。符号を定義するメタデータは、より一般的なメタデータが、規準と手引きのデータベースに記録されているのにたいして、検索表として与えられている。

第4章 一般的な問題と解決策

4.1 はじめに

第2章では、行政的なデータソースを利用するための多くの妥当な理由を概説したが、その利用に関係した多くの問題もある。それらの問題のいくつかは特定のデータソースや特定の利用に固有のものであるが、多くは、事実上、一般的なものだろう。本章では、いくつかの一般的な問題を概説し、それらを解決するための方法、あるいは少なくともそれらの影響を最小化するための方法を提案する。第1の行政的なデータソースの利用に特有な問題、そしてデータ連結に特有な問題は、第3章と第6章において、それぞれ別々に取り扱う。

4.2 世論

第2章では、いくつかの国における世論がデータ共有をいかに支持するのか、ということ考察した。しかしながら、データが政府のいたるところで共有されることに、国民が不安を覚える国もある。そのような懸念を減らすことはきわめて難しいが、考えられる方法には、データ利用について、明確な制限や規則を公表することである。この公表によって、国民および企業に、統計目的のために利用され、または収集される機密性を要するデータが譲渡元の政府機関（とくに課税や各種給付に係る機関）に戻されないと理解されるようにする。

これは国連「政府統計の基本原則」と整合的であって、その原則5（「統計目的のためのデータは、統計調査であれ、行政記録であれ、あらゆる種類のデータソースから獲得することができる。統計機関は、精度、適時性、経費と回答者の負担を考慮することによってデータソースを選択するべきである。」）は行政的なデータの利用を促進する。原則6（「統計の編成のために統計機関によって収集される個別データは、自然人にかかわるものであれ、法人にかかわるものであれ、厳密に機密扱いとし、もっぱら統計目的のためだけに利用されるべきである。」）と合わせて、原則5はデータの一方向の流れの原則を確立する。

データ共有に否定的な世論を克服するのに役立つ他の方法には、政府と回答者の双方にたいする、さまざまなデータソース利用にかんする費用・便益分析の公表がある。行政的なデータソースを利用するとき、マイクロデータがいっそう安全であると主張することもできるだろう。どの調査票も郵送されないし、調査員がデータを紙に書いて、あるいは電子情報として、保持することもない。統計作成過程に必要な事務職員が少なくなるので、機密性を要するデータを取得できる人も少なくなる。

4.3 国民の関心度

調査を通じた国民との直接的な接触は統計機関にたいする関心度を高めるために役立つ。行政的なデータの利用はそのような接触を減らすことができるし、統計機関の活動にかんする国民の認識を低下させることもできる。もし、それが問題になるとときには、もっとも明らかな解決策は統計機関とそのデータ作成にかんする「マーケティング」を改善することである。これは、行政的なデータソース

の利用による節約のごく一部を、マーケティング予算に振り替えるように求めることがある。

おそらく国家の統計機関の活動および統計作成を促進するもっとも効果的な方法 — とくに中期から長期において — は、教育機関、企業集団およびその他の目標となる顧客層との強力な連携を確保することである。利用者集団もまた、このような連携の確保という点で、とくに重要であり、積極的に働きかけるべきである。

4.4 変化の管理

公共部門の行政的なデータソースは一般に税金を徴収するため、あるいは政府の政策を監視するために設定されている。これは、行政的なデータソースが政策の変更に影響を受けやすいことを意味する。もし政策が変更されると、行政的なデータソースは捕捉範囲、定義や閾値などにかんして、影響を受けることがある。あるいは、完全に廃止されることさえあるだろう。行政的なデータを保持・処理するために利用されるコンピュータシステムの変化もまた、統計目的のためにデータを供給することに影響を与えることもある。民間部門のデータソースでさえも、こうした種類の変化から影響を受けないことはない。もっともこの場合、データソースの変化は、市場要因の変化によって推進される可能性が高いが。

そのような変化は警告なしに突然起こるかもしれない。とくに、政府の変化、大臣の交代ないし法律の変化の直後は危険性が高い時期になる傾向がある。数年前にスロベニアからつぎのような例が報告された。そこでは雇用にかんする行政的なデータの供給が、大臣の交代後、しばらくの間、停止され、雇用統計の作成にかんして、統計局に深刻な問題をもたらした。その後、この種の変化の可能性や影響を最小化するために、法律に裏打ちされた手続きが実施されている。

したがって、特定のデータソースに依存することは、いつでもある程度の危険性を帯びる。この危険性はある程度まで、法的条項あるいは契約条項によって処理することができる。実際、そのような問題を避けるための最善の方法は、行政的なデータソースの責任者と定期的な接触をとる方向に向かう。これは、その責任者に統計にたいする要求を認識させるようにするためであり、あらゆる可能性のある変化にも影響を与え、その兆候を早期に得るためである。特定のデータソースに強く依存するところで、もしそのソースが利用できないならば、何ができるかを定める緊急対応策を準備する価値もある。事後的に対応をとらざるをえないよりも、事前対策を講じるほうが明らかに良い。

4.5 単位

行政的なデータソースを利用するとき、しばしば直面する1つの大きな問題は、そのソースでもちいられる単位が、要求される統計単位の定義に直接的に一致しないことである。行政単位（法律的单位、課税単位、請求者など）から統計単位（企業、人、世帯など）へ変換する過程は、概念的にきわめて難しいところがあるし、しばしばなんらかのモデル化にかかわっている。

企業統計において、この過程はプロファイリングとして知られており、一般的に企業統計登録簿の機能である。欧州連合統計局は、企業登録簿勧告マニュアル¹⁵⁾の第19章で、この過程についてのガイ

ドラインを公表している。そこでは、プロファイリングが「企業集団内の統計単位、その連結、および統計データ収集にもっとも効率的な構造を確立するために、国家的な次元や世界的な次元で、企業集団の法律的、経営的および会計的な構造を分析するための方法である」と定義されている。

図4.1は、一組の連結している企業単位の構造が、統計の観点と比較して、法律的不いし行政的な観点から、いかに大きく違って見えるかを示している。プロファイリングは、上で定義されたように、統計構造を生み出し、それを法律的不いし行政的な構造に写像する過程と見ることができる。

プロファイリングは、複雑な単位構造にかんして、より良い理解を与えるとはいえ、それは経費も時間もかかり、訓練された職員を必要とする。したがって、ある経済のすべての企業単位について、職員による詳細なプロファイリングを試みることは全く非現実的である。最大の利益を与えるような事例に焦点を合わせることが必要である。プロファイリングは以下3つの要因間のトレードオフあるいは妥協と見ることができる。

- ・分析される企業構造の数量
- ・プロファイリング分析の精度ないし深さ
- ・利用可能な資源（経費と職員の適合性の両者によって決定される）

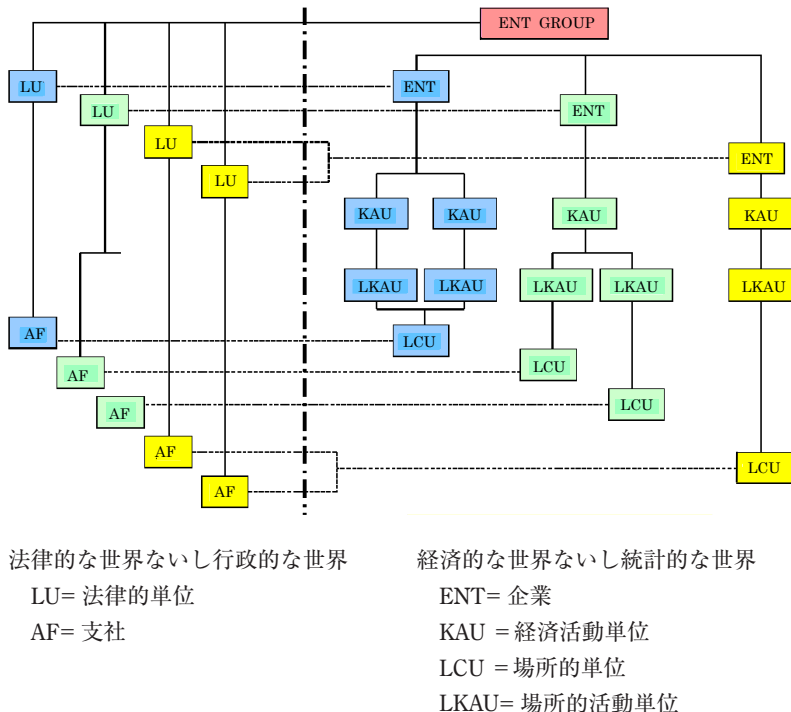


図4.1 企業集団の各単位にかんする異なった観点

15) 欧州連合統計局の以下の URL で入手可能である。

http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-BG-03-001/EN/KS-BG-03-001-EN.PDF

囲み資料4.1は、欧州連合内の企業に統計的定義を適用する整合性にかんする調査研究の一部として、3ヶ国（デンマーク、オランダ、イギリス）において、それぞれ独立にプロファイルされた企業構造の4つの例を与えており¹⁶⁾、つぎのことを明らかに示している。プロファイリングはある程度、芸術的であること、しかし、その特定の課題がプロファイリング規則を調和させるためのかなり方法的な作業 — これは上記の欧州連合統計局・企業登録簿勧告マニュアルの第19章で部分的に示されている — をもたらしたとはいえ、必ずしも「正しい」答は存在しないことである。

職員によるプロファイリングは、大きな母集団のすべての単位については非現実的であるが、ある形式の自動化された、規則にもとづくプロファイリングが存在することがある。単位間における連結関係の属性ないし性質にもとづく標準的な規則は、多数の統計分野における、行政単位と統計単位との間の相違を克服するのに役に立つ。たとえば、統計上の世帯は建物に居住する個人と個人との関係にもとづいて導出することができる。この方法は北欧諸国において適用されている登録簿にもとづく人口センサスの方法論において首尾よく使われている。

いくつかの事例で実効性があるプロファイリングの代替手段は、統計的な「調整」を行うことによって、単位の定義上の相違について修正を検討することである。この方法のおおざっぱな例は、統計単位が人、行政単位が仕事の場合であろう。労働者が平均して1.15の仕事をもつと仮定すると、この調整因子を仕事数から雇用者数を推計するために利用することができる。

囲み資料 4.1 プロファイリングの演習：企業はいくつか

下記の例は、「企業概念の異なった解釈の影響」という調査研究 — これは欧州連合統計局のために、デンマークおよびイギリスの情報提供をうけ、オランダ中央統計局が準備した調査研究である — から引用されている。各例の後に、参加3ヶ国のそれぞれによって与えられた回答がその根拠の要約とともに示される。各例はつぎのような企業の定義にもとづいている。

「**法律的単位の最小の結合体であって、財ないしサービスを生産する組織単位である。そして、意思決定におけるある程度の自律性から利益を得る。……企業は単一の法律的単位であることがある。**」

出所：EUの統計単位にかんする規約第696/93号

例1 企業集団の2つの法律的単位にはNACE（欧州共同体の経済活動一般分類……訳者注）の4桁の分類コードがある。2つの法律的単位は主として集団外の第三者に販売している。それらは建物、経営、購入物および雇用者を共有する。

回答

- ・オランダとイギリス：共有される生産要素の集中度を考慮して、1つの企業にまとめる。

16) 「企業概念の異なった解釈の影響」 — これは欧州連合統計局のために、デンマークおよびイギリスの情報提供をうけ、オランダ中央統計局が準備した調査研究である。

・デンマーク：2つの法的単位とも、集団外に50%以上を販売しているので、独立した2つの企業である。

例2 4つの法的単位：AとBには違った活動があり、共同購入はしていないが、建物は共有する。CとDは建物、雇用者および購入物を共有する。すべての4つの法的単位が1つの会社と自称する。

回答

- ・オランダとデンマーク：AとBは独立した企業であり、CとDを組合せて1つの企業とする。なぜならば、AとBが市場条件のもとで操業する一方で、CとDは生産要素を共有するからである。
- ・イギリス：4つのすべての法的単位を組合せて1つの企業とする。なぜならば、それらの法的単位は1つの会社と自称するからである。

例3 3つの法的単位：そのすべてが主として外部の消費者向けに生産し、経営と購入物を共有し、1つの会社と自称する。AとBは建物を共有する。BとCは同じ活動を行い、雇用者と資本財を共有するが、独立したデータを供給できない。

回答

- ・オランダ：組合せて1つの企業とする。なぜならば、すべての法的単位が経営と購入物を共有し、1つの会社と自称するからである。
- ・イギリスとデンマーク：BとCを組合せて1つの企業とする。なぜならば、それらは水平的に統合されており、2つを一緒にしたデータのみが利用できるからである。Aは独立した企業である。

例4 12の法的単位が企業集団を形成する。1つの法的単位のみが活動しており、その他は雇用者がいない。

回答

- ・オランダ：活動している単位のみから成る1つの企業である。なぜならば、活動していない単位は企業の一部ではないからである。
- ・イギリス：すべての単位から成る1つの企業である。なぜならば、活動していない単位について、独立した企業があるとするのは意味がないからである。
- ・デンマーク：各単位はそれぞれ独立した企業である。なぜならば、単位間に強力な連携が存在しないからである。

4.6 変数の定義

単位の定義の相違だけでなく、行政システムと統計システムの間には変数の定義の相違もありそうである。行政的なソースのデータは一般に、特定の行政目的のために、収集されており、その目的に関係する要求と優先度は統計システムの場合と違いがあるかもしれない。たとえば、付加価値税（VAT）が適用される売上高には、それが適用されない財・サービスの販売に関連した売上高を含まないことがあるのに、統計システムでは総売上高が適当である。

もう1つの一般的な例は失業の定義である。その標準的な統計の定義はつぎのとおりである¹⁷⁾。

『『失業者』には該当期間中に以下であった規定年齢以上のすべての人を含む。

(a) 『仕事をもたない』、すなわち、有給雇用でも自営業でもなかった。

(b) 『現在、就業可能である』、すなわち、該当期間中に、有給雇用ないし自営業に、就くことができた。さらには、

(c) 『仕事を探していた』、すなわち、有給雇用ないし自営業を探すために、特定の最新期間に具体的な措置を講じていた。』

しかしながら、行政的なデータソースにおける失業の定義は、たいてい、失業手当の請求者数ないし求職登録者数にもとづいている。失業している人々の中には、すぐに仕事をみつけることができると思って、失業者として登録しない人もいるかもしれない。いくつかの文化では、失業手当の請求にともなう社会的な不名誉が存在するかもしれない。他方、失業手当の請求者の中には、就業できない人や積極的に仕事を探さない人もいるかもしれないので、統計的に、失業者として数えられるべきではない。

定義が異なる問題を解決するための第一歩は、その相違を理解し、その影響を数量化しようとすることである。いくつかの相違は、実際には、実質的な影響を与えないだろうから、無視しても差し支えない。その他の相違は系統的なものであるだろうから、データの調整によって解決することができるだろう。ときどき、さまざまなデータソース — とくに上述した売上高の例のような財務会計的な変数のためのソース — の変数を組合せることによって、その相違の影響を導出したり、推計することができるだろう。いくつかの事例では、行政的な定義に影響を与えることさえできるだろう。

4.7 分類体系

変数の場合と同様に、行政的なデータソースのなかで利用される分類体系は、統計的な世界で利用されるそれとは異なることがある。たとえ両者が同じだとしても、前者は、行政的なデータソースの主な目的に応じて、おそらく単位の特定の属性に焦点を合わせることによって、後者とは異なる仕方でも適用されるだろう。たとえば、許認可、保健・安全ないし環境保護に関連した行政的なデータソー

17) 第13回国際労働統計家会議（1982年10月）によって採択された、経済活動人口、雇用、失業および不完全雇用にかんする統計についての決議を参照せよ。

http://www.ilo.org/global/statistics-and-databases/standards-andguidelines/resolutions-adopted-by-international-conferences-of-labourstatisticians/WCMS_087481/lang-en/index.htm

スは、統計目的のために必要な企業の主な経済活動ではなくむしろ、そのソースにもっともかかわる企業の経済活動に関心があるだろう。

他の事例では、行政的なデータソースの分類は統計目的のために必要な詳細レベルで適用されないこともある。あるいは、分類は単に、行政的なデータソースにとって重要度の高くない変数であろう。これは精度の不足という結果になるだろう。

分類の体系ないし版が異なる場合、通常の解決策は行政分類の分類コードを統計分類のそれに対応づけるために変換行列を作成することである。そのような対応づけは1対1、多数対1、1対多数ないし多数対多数であるだろう。後2者の場合では、ある種の確率的な割り当てが求められるだろう。

囲み資料 4.2 単純な変換行列の利用

分類コード1	分類コード2	ウェイト	
0100	01300	100	1対1の相互関係
0101	01210	26	1対多数の相互関係
0101	01221	14	
0101	01222	29	
0101	25730	11	
0101	74332	20	
0102	03200	100	
0103	01300	36	
0103	74332	64	

この表は、変換行列全体からの一部の引用であって、1つの分類体系からもう1つのそれに変換するとき、見受けられる主な問題を例示している。この場合、行政的なデータソースでもちいられる分類コード（分類コード1）は、ウェイトにもとづく確率的方法で、統計システムでもちいられる分類コード（分類コード2）に対応づけられる。

したがって第1の問題はそのウェイトを決定する方法である。ウェイトを推計することはできるが、望ましい方法は、可能であれば、いくつかの分類コードの組合せをとともう各単位の割合を検討することによって、2つのシステムに応じて分類される単位の分析からウェイトを導出することである。符号化の誤りの影響を減らすために、有効ないし妥当とみなされる分類コードの組合せのみに及ぶように、単位の分析を制約することが必要であるだろう。

引用された変換行列の第1行は、100%のウェイトによって表示される1対1の相互関係を示している。これは、行政の分類コード0100をもつすべての単位が統計の分類コード01300に割り当てられるべきであることを意味する。つぎの5つの行は、1対多数の相互関係を示す。もし単位が行政の分類コード0101をもつならば、統計の5つの分類コードに入る可能性がある。これら統

計の分類コードのそれぞれについて、単位が正確な分類コードに入る可能性がウェイトに反映される。こうして、01210が正確な統計の分類コードである見込みは26%である。

この場合、行政の分類コード0101をもつ単位に正確な統計の分類コードが与えられる確率は、それぞれの組合せの確率を2乗和することによって、つぎのように計算することができる。

$$0.26^2+0.14^2+0.29^2+0.11^2+0.2^2=0.2234$$

これは、行政の分類コード0101をもつ単位が、誤った統計の分類コードを与えられる見込みが77.66%であることを意味する。この可能性は受け入れがたいほど高いと思われるかもしれないが、たとえ付与された分類コードが単位レベルで誤っているとしても、ウェイトが正確であれば、分類コード間の単位の分布は集計レベルでは正しいはずであって、変換行列の適用に系統的な偏りがない限り、このような方法で符号化された単位の統計データには、偏りが生じないはずであることを忘れてはならない。

上記の引用例のような変換行列は、一方向的であることにも留意すべきである。さまざまなウェイトがある独立した行列は、統計の分類コードから行政のそれに変換することが求められるだろう。たとえば、一方向における1対1の相互関係は、逆方向における1対多数の相互関係となることがある。これは上記の表に例示されている。ここでは、分類コード0100と01300の間で1対1の相互関係があるが、統計の分類コードから行政のそれに変換されるとき、01300は0100ないし0103に対応づけることができよう。

正確性がマイクロデータレベルで求められる場合では、変換行列による方法に重大な限界があることは、囲み資料4.2から明らかである。いろいろな他の方法の利用が資源およびデータの利用可能性に応じて考えられるかもしれないが、有効な第一歩は、分類データが行政的なデータソースによって、どのように収集・処理されたのかについて、さらには分類データが利用される行政活動の性質について、詳細な理解を必ず得ることである。

いくつかの事例では、正確な統計の分類コードを選択する可能性を高めるために、利用できる他の変数が、行政的なデータソースのなかで利用できることがある。そのような1つの変数は、行政の分類コードが導出されたテキストの記述である。これが利用できれば、統計家にとって、行政の分類コードそれ自体よりも役立つ可能性がある。なぜならば、統計家はテキストの記述から直接的に正確な統計の分類コードを導出するために、手作業による技法ないし自動的な技法を適用できるからである。この方法は、テキストの記述が行政の分類コードと統計のそれとの間で1対1の相互関係が存在しない場合のみ符号化されるので、変換行列法と組合わせて利用することができる。もっとも、もし符号化の精度が行政システムと統計システムの間で異なれば、潜在的な偏りの危険性があるが。

いくつかの国でうまく利用されている1つの方法は、統計システムと行政システムの双方で利用するための自動的な符号化の手段を開発することである。これは符号化の高度の整合性を保証し、共通の分類体系を利用することを強力に推進する（だが、必ずしもその利用を強制しない）。

双方のシステムで利用することができる符号化の手段に加えて、符号化の専門技術と行政的なデータ供給者の研修にかんする規定は、符号化の整合性を改善するのに役立つ。同時に、統計家が共通の分類体系がそなえる利点を強調することは、必ず有益である。それはまた、分類体系の改訂を早期に通告することに役立つし、もたらされた変化が現存している間に、行政的なデータの供給者にできるだけ多くの支援を与えることにも役立つ。

4.8 適時性

統計目的のために使われる行政的なデータの有効性に影響を与える適時性にかんして、以下のよう
な3つの独立した問題がある。

- ・行政的なデータが、統計にたいする要請に応じるには、時間的に遅れるために、利用できないことがある。
- ・行政的なデータは、統計の該当する期間と一致しない期間に関係していることがある。
- ・行政的なデータは、ある期間にわたって測定されることがあるのに、統計の要件はある特定の時点である（あるいは逆も同様）。

最初の問題を考えると、事象が現実の世界で発生し、それが行政的なデータソースによって記録される間に、ある種の遅れ (lag) が一般に存在するだろう。それから国家の統計機関がデータを利用することができるまでに、遅れの拡大が続く。以下の図4.2は、イギリスにおいて、企業が活動を始め、それが企業統計登録簿に記録されるまでの全体的な遅れの日数を示している。企業の生・死に関する遅れは、企業登録簿の捕捉誤差 (coverage error) の主な原因である。それらの遅れを測定するならば、登録簿データにもとづくどの統計においても、遅れにたいする許容幅を設定することができる。

このようにして遅れを分析することによって、その影響を推計するための簡易な統計を作成することができる。たとえば、上記の例において、2/3の企業は活動開始から2ヶ月以内に企業統計登録簿に現れる。遅れの平均は約120日であるが、その分布のきわめて長い末尾の外れ値 (図4.2では省略されているが、最長10年の遅れに関係した、もっとも極端な場合のような) によって影響されるので、その数字がとりたてて有用というわけではない。おそらく、この場合の有効な平均の尺度は中央値であって、それでは約40日である。この分析のもう1つの興味深い特徴は、少数のマイナスの遅れである。これは企業が事業開始のずっと前に、登録手続きを完了するとき起こりうる。

このような分析は、統計家が統計を編成するために利用するデータソースの遅れの性質や影響を理解するのに役立つために、重要なことは明らかである。それはまた、統計の精度を改善するための調整を行うのに利用できる情報も与える。

遅れの存在と長さは、短期統計について、行政的なデータソースの利用を困難にすることがある。たとえば、6ヶ月間の遅れは重要な月次の経済データ系列のためにはおそらく受け入れがたいであろうが、年次統計については、たいした問題ではないだろう。

遅れの問題を解決するための第一歩は、上記のような分析を用意することによって、その影響を理

解することである。ひと度、遅れの影響を理解すると、それを調整するためのモデルを開発することができるだろう¹⁸⁾。反対方向の遅れが相互に打ち消し合う比較的に安定的な系列にあてはまるだろう。たとえば、図4.2の企業登録の遅れは、企業母集団にかんするデータを作成するために、登録抹消の遅れによって相殺されるだろう。しかしながら、これが経験的証拠なしに当てはまると想定することは危険でありうる。

遅れの性質と影響が決定されると、遅れを引き起すものが何かを理解しようとすることは有益である。いくつかの事例では、遅れを小さくするような行政的なデータソースの変更を提案することができるだろう。これは、統計家と行政的なデータソースの双方にとって有益であるだろう。

適時性に関係する第2の問題は期間の相違である。たとえば、年度納税申告書のデータはしばしば、税年度末の数ヶ月後に入手できるにすぎないから、おそらく月次統計ないし四半期別統計には不適切である。しかしながらいくつかの事例では、年次の行政的なデータを短期統計に利用することができる。そのデータが年次のベースで収集されているならば、とくにそうである。これは、行政的なソースによるデータの収集・処理にかんする作業総量を1年中に分散する要請があれば起こりうる。年を通じて収集されるデータの単位の発生が十分に不揃いであるかぎり、そのようなソースから有意義な

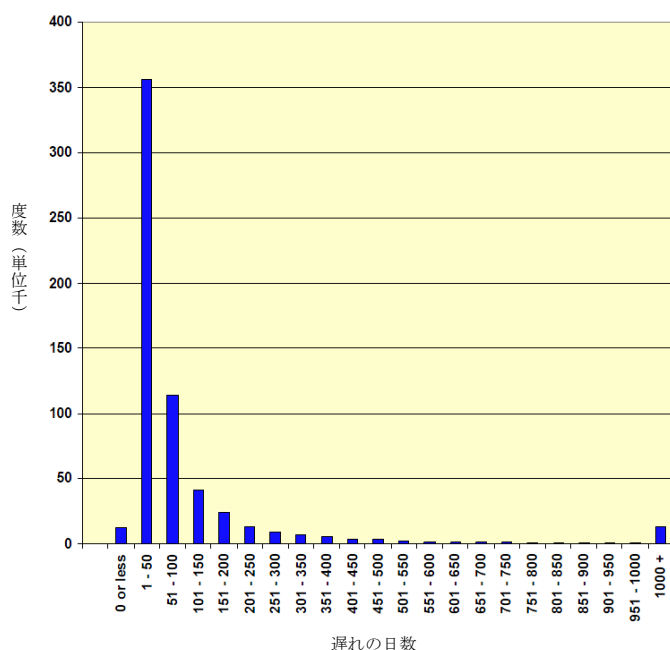


図4.2 イギリスにおける企業登録の遅れ¹⁹⁾

18) 図4.2に関係した例については、Annex B of Business start-ups and closures: VAT registrations and de-registrations in 2005 -Guidance and Methodology を参照せよ。 <http://stats.berr.gov.uk/smes/vat/VATGuidance2005.pdf>

19) 出所：Model Quality Report in Business Statistics, Volume III, Eurostat。 <http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/MODEL%20QUALITY%20REPORT%20VOL%203.pdf>

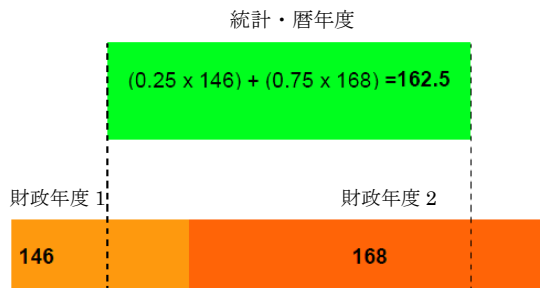


図4.3 期間の食い違いの処理

月次ないし四半期の統計的なトレンドデータを導出することができるだろう。

図4.3は、行政的なデータが4月1日から3月31日の財政年度にもとづくのにたいして、統計の要件が暦年データであるような場合を示している。行政的なデータを変換するためのもっとも単純な方法は、第1の財政年度から数値の25%、第2の財政年度から数値の75%を合計することである。この方法は、もしデータの長期的な傾向が適度に安定的ならば、妥当な近似値を与えるはずである。もっとも、より変化しやすい系列については、他の複雑な推計方法が必要とされるが。

第3の問題は、特定の時点にかかわるデータと期間にかかわるデータ（すなわち、年次平均ないし月次平均）との間の相違に関する。たとえば、特定の期日にかかわる雇用データにたいする統計的な要求が存在するだろう。これにたいして、行政的なデータは月次平均を与えるにすぎないだろう。

前の例と同様に、解決の第一歩はその相違の影響を分析することであり、その影響がさらに別の作業を必要とするほど十分な意義があるかどうかを決定することである。1つの考えられる解決策はモデルベースの数理的な調整である。たとえば、統計の期日が月初め頃であれば、前期の平均値を考慮に入れるモデルが適切であるだろう。代替的な方法は、行政的なデータを調整するために比較的の小規模な調査の結果を利用することだろう。

4.9 データソース間の非整合性

多数のデータソースを利用するときの1つの特別な問題は、それらソース間の非整合性に関する。1つのソースのデータは、別のソースのデータと矛盾するように見えるかもしれない。これは異なる定義や分類、タイミングの相違によることがあるし、単に1つのソースにおける誤りによることもある。この問題は行政的なデータを統計データと比較するとき、あるいは2つの行政的なデータソース（または2つの統計データソース）を比較するとき、起こりうる。

そのような齟齬を解決するために、どのデータソースが特定の変数にたいして、もっとも信頼性が高いかを決定することによって、優先規則を設けることが必要である。ひと度、データソースの優先順位がある変数にたいして決定されると、それから優先度の高いソースのデータは優先度の低いソースによって上書きされないことを確保できるはずである。この過程は、いくつかの利用可能なデータソースの変数とともに、ソースコードが保持されていれば、ずっと容易に処理される。日付の利用と

保持もまた役に立つ。1つのデータソースがもう1つのものよりも信頼できると考えられる時でさえ、10年前の信頼できるソースのデータはさほど信頼できないソースの最新期間データよりも精度が低いかもしれないからである。いくつかの事例で適切な単純な方法は、高い精度のデータを低い精度のデータに上書きできるようにしながら、優先順位の逆にデータを読む込むことである。

たいていの場合、いくつかの関心のある変数があるだろうし、優先度は変数によって異なるかもしれない。たとえば、労働者雇用にかんする行政的なデータソースは（法定）雇用の妥当な推計値を与えるだろう。雇用変数とそのソースの中心的な機能に密接に関係するからである。しかしながら、それは、雇用主の経済活動を決定するにはあまり良くないだろう。雇用主の経済活動はそのソースの目的にたいして、副次的に重要であるにすぎないからである。こうして、多数のデータソースが雇用データを導出するために利用されるならば、最適な統計データセットを導くために、各データソースにおける各変数の相対精度を考慮に入れることが必要であるだろう。

利用されるデータソースが多いほど、この比較過程はますます複雑になるが、多数のデータソースが利用できることはしばしば、データの精度を確認するために役立つ。いくつかの事例では、一定のデータソースは統計作成に直接的に利用されるのではなく、精度保証過程の一部として、評価基準を検討するためだけに利用されることがある²⁰⁾。さまざまなデータソースの精度について、もたらされる認識はまた、当該のデータソースに返されることができし（通常、統計の機密性を守るために、単位レベルよりも集計レベルで）、そのソースの精度の改善にかんする議論の基礎を提供することもできる。

囲み資料 4.3 ソースが異なるデータ

	データソース 1	データソース 2
	教育登録簿	住民登録簿
氏名	Steve Vale	Stephen Vale
住所 1	5 St Peter's St	5 Saint Peters Street
住所 2	Machen	Machen
住所 3	Newport	Caerphilly
住所 4	Gwent	South Wales
郵便番号	NP1 8QB	CF83 8QB
生年月日	28/12/1967	28/12/1997
職業	Statistician（統計家）	Civil Servant（公務員）
雇用主	CSO（中央統計局）	Office for National Statistics（国家統計局）

20) 評価基準の検討例で、企業統計登録簿の捕捉範囲を職業別電話帳のそれと比較するために地図を使用するものは、つぎの論文に見出すことができる。

“The Development of Small-area Business Statistics in the United Kingdom” at <http://live.unece.org/fileadmin/DAM/stats/documents/ces/sem.53/wp.7.e.pdf>

職場の郵便番号 NP10 9XX

NP10 8XG

この例は、著者にかんする架空データを含む2つのレコードを示している（イギリスでは、教育登録簿と住民登録簿はまだ存在しない〔既に存在する……訳者注〕）。それはソースが異なるデータを調整しようとするときに、いくつかの一般的な問題を例示することを意図している。

- ・ 誤り — 簡単な妥当性の審査は住民登録簿の誤りを見つけることである。1997年生まれの人はまだ就学中であろう。そのために職業や雇用主をまだもつことができないだろう。教育登録簿では出生年が1967年であるので、これは単なる入力ミスのように見える。自動的な審査は通常、そのような疑う余地がない誤りを見つけることができる。とはいっても注意して使用されなければならないが。たとえば、養子が原因で、両親よりも高齢の子供たちにかんする少数の事実の事例がフィンランドで検出されている。
- ・ タイミング — 与えられた住所と郵便番号は時点が異なるが、実際に、同じ建物を指しているかもしれない。その相違は郵便区域の境界変更が原因かもしれない。これは、過去の住所ファイルを調べることによって、あるいは地理情報システムを利用し、現在の住所と過去の住所を対応づけることによって確定することができよう。
- ・ 略語 — 教育登録簿における住所1の文字列の終わりにある“St”は“Street”の一般的な略語であるから、“St”がテキスト文字列の終わりに現れるとき、同義語として処理されるべきである。しかしながら、つぎのことに留意すべきである。すなわち、住所1の文字列の最初にある“St”は“Saint”の略語としても使われているので、注意をいくらか払うことが必要である。同様の事例を他の用語においても見出すことができる。
- ・ タイミングと略語 — イギリスにおける“CSO”とは、“Office for National Statistics（国家統計局）”の旧名“Central Statistical Office（中央統計局）”の略語であり、その変更をよく知らない人々が、いまなお使用しているおそれがある。
- ・ 綴りの違い — “Steve”と“Stephen”は同名の異綴語であり、そのようなものとして処理されるべきである。
- ・ 分類の問題 — 職業の“Statistician（統計家）”と“Civil Servant（公務員）”は相互に排他的ではない。“Statistician”は専門職を指すといえよう。これにたいして、“Civil Servant”はむしろ雇用の性質に関係している。
- ・ 欠損値 — とくどき、値が欠けていたり、部分的にのみ存在しているとき、何らかの欠損値が使われる。一般的な値は“Z”ないし“999999”である。イギリスにおいては、郵便番号の後半部が分からないときには、「職場の郵便番号」の項目で見られるように、欠損“9XX”がしばしば使われた。残念ながら、この欠損値の使用は、郵便局が“9XX”で終わる、実際の郵便番号を配布し始めたときに、放棄されねばならなかった。

4.10 欠測値

欠測値の問題は行政的なデータソースにかぎらない。それは、また統計調査にたいする全面的ないし部分的な非回答によることがある。あるいは、編成処理（editing process）におけるデータ値の除去さえも原因となる。しかしながら行政的なデータソースでは、とくに欠測値の問題がしばしば系統的なものであるために、論点がときどき異なることがある。

この主な理由は、特定の変数が行政的なデータソースによって、まったく収集されないことがあるからである。あるいは、固有の行政的な要求があるときに、一定の単位から成る分類項目についてのみ、特定の変数が収集されることがあるからである。変数はまた単に、行政目的にとって重要度が低いこともあるから、そのソースの所有者は欠測値を問題視することはない。

統計調査における非回答を処理するための標準的な解決策のいくつかはまた、行政的なデータソースにおける欠測値の問題を解決するために利用することができる。推論的な「ホットデッキ（hot-deck）」代入法ないし「コールドデッキ（cold-deck）」代入法のようなさまざまな欠測値の補完方法は、この問題が単位の一部に影響を与えるにすぎない場合、しばしば適切である。ほとんど、あるいは全部の単位が影響をうける場合、モデル化の方法が適切であろう。

囲み資料 4.4 事例研究：行政的なデータの欠測値の処理 — 1人当たり売上高比率

企業規模を測定するためにもっとも一般的に利用することができる2つの変数は、雇用者数と総販売額（売上高）である。しかしながら、それらの変数の一方ないし双方が新企業、とくに小企業について、欠けていたり、信頼できないのが一般的である。

この問題の解決に役立てるために、1人当たり売上高比率を欠測値の推計に利用することができる。その比率は、双方の変数が存在し、信頼できるとみなされる類似企業の情報を使って、経済活動および制度部門にもとづくさまざまな分類項目における平均的な1人当たり売上高比率を計算することで作成される。

たとえば、以下は、国際標準産業分類（ISIC）のさまざまな階級について計算されたダミーの1人当たり売上高（TPH）の値である。

ISIC の階級	TPH
.....	
45.11	95
45.12	68
45.21	149
.....	

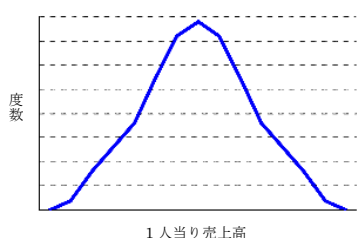
もし、ISIC の階級45.12に属する企業の売上高が200であり、雇用が欠けているならば、補完される雇用値はつぎのとおりである。

$200/68 = 2.94$ (3に丸められる)

1人当り売上高の値を計算するとき、外れ値の問題にしばしば直面する。それで刈り込み(値の上位および下位のx%を除去)や、四分位範囲にかんする平均を計算するような方法がしばしば使われる。

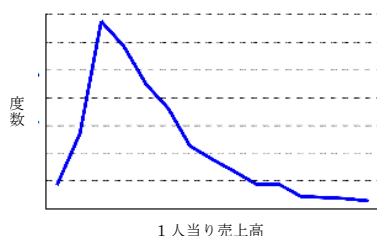
この種の比率はまた、更新を有効に行うため、ソースが異なるレコードを照合するため、さらには誤りを見つけるために、一般的に利用することができる。たとえば、1人当り売上高の値の分布をグラフ化し、調べることによって、しばしば問題となっている単位の有効な母集団情報を得ることができる。つぎの図は、そのような演習問題の観察例である。

1. 正規分布



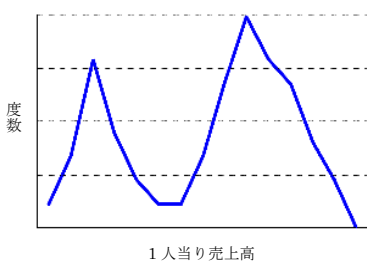
この場合、1人当り売上高の値は、平均値のまわりに均等に分布している。これは、母集団における単位の均質さを相対的に示し、外れ値の影響がきわめて限られることを示している。

2. 非対称分布



この場合、比較的小さい値のまわりの単位をグループ化する必要性は明らかであって、右側末尾の外端近くの外れ値は分布の平均に影響を与える。これは1人当り売上高の比較的一般的な分布であり、外れ値の影響を減らすための措置をとる必要性を浮き彫りにしている。

3. 双峰分布



この場合は、問題となっている母集団がむしろ不均質であって、より有意義な1人当り売上高比率を得るために、2つの部分母集団に分割する価値があるだろうということを例示している。

4.11 変化への抵抗

政府統計において、行政的なデータソースをいっそう効果的に利用するための主な障壁の1つで、もっとも認識されていないものの1つは、組織内部から起こる。統計家は行政的なデータの利用に抵

抗することがある。なぜならば、彼らは、自身で収集しないデータを信頼しないからである。彼らはしばしば、行政的なデータの否定的な特質の側面ばかり注目する。そして、調査回答が実際に統計的基準に適合するという、ほとんど検証されない前提にしばしばもつづいて、調査データの精度には楽観的に過ぎる見解をもつ。

解決策は明らかに、行政的なデータソースが与える可能性について、統計家をより良く教育することである。これによって、統計家はあらゆる次元の精度にかんする幅広い見識をもち、データ供給者とデータ利用者にたいする影響に焦点を合わせるように促される。これに関連して、調査データと行政的なデータにかんする実質的な相対精度を決定することが重要である。たとえば、行政的なソースのデータは統計的な定義の要件に対応しないのに、政府の調査データは対応するとしてしばしば考えられている。しかしながら、實際上、実質的な相違は存在しないかもしれない。とくに、統計調査への回答者が、特定の変数が統計目的のために定義される方法について、しばしば長々とした注記を読まないで、最新の行政申告書から数値を単に転記するならば、そのような相違は存在しないであろう。

内部の抵抗という障壁を乗り越えるのに役立つ追加的な方法は、行政的なデータの利用による経費節約が必ずしも職員の削減を意味しないことを示すことである。節約された資源を、少なくとも部分的に、統計の精度を改善するために、あるいは統計の範囲ないし頻度を増すために利用することができる。

4.12 要約

本章は、行政的なデータソースを利用するときに克服すべき多くの問題があることを明確に示している。本章の狙いは、他の章でも、それらの問題に直面することを示し、たいていの場合、完全ないし部分的な解決策を見つけることができることを示すことである。本章では、読者が直面するかもしれないすべての問題を取扱うことはできていない。データソースに固有の問題はとくにそうであるが、本章の意図は、特定の状況に対応するために適用することができる着想を与えることにある。

全体的にみて、多くの他の統計分野と同様に、効果的な計画立案や経営、データソースにかんする十分な知識、創造的な思考、さらには経験を交換し、他の経験から学ぼうとする意欲によって、統計目的のために行政的なデータを利用するときに直面するたいていの問題を克服することができるし、少なくとも減らすことができると言って間違いない。

行政的なデータは、しばしば統計データソースと異なる処理を要求する。統計の作成過程を変更せずに、統計データの代わりに行政的なデータを使うだけでは、實際上、めったにうまくいかないであろう。

記憶にとどめるべき最後のことは、あらゆる問題にもかかわらず、行政的なデータを利用する利益は、それでもなおしばしば、経費よりもずっと大きいことである。

浜砂 敬郎〔九州大学名誉教授〕

西村 善博〔大分大学経済学部 教授〕