



自治体行政データ

田中 隆一
(東京大学教授)

1 自治体行政データとは？

日本では、誕生したほとんどの人々について出生届が作成され、生まれた日を含めて14日以内に区市町村役場に提出される。出生届には名前とともに「生まれたとき」、「生まれたところ」、住所が記載される。さらに、生まれた子の父と母の情報として氏名、生年月日、本籍などが記載される。これらの情報をもとに、新たに生まれた子は、戸籍と住民基本台帳へ記録されることになる。

役場では、子の健康保険への加入や乳幼児医療助成・児童手当金の給付を受けるための手続きも行う。健康保険のうち、国民健康保険の窓口は区市町村自治体である。国民健康保険に加入している人々の医療費の大半は保険でカバーされ、その額を決めるのに必要な診療履歴が自治体に集まる。乳幼児医療助成制度は自治体に居住する乳幼児が医療機関で受診した時の保険診療自己負担分を居住自治体が負担する制度である。助成を受けると、国民健康保険と同様に受診情報が医療機関から自治体に送られる。さらに、児童手当金は0歳から中学校卒業までの児童を養育している人々に給付されるが、給付額の決定に所得制限限度額がある場合には、家計所得の情報と照らし合わせる必要があり、親の所得情報が自治体に集まる。

子どもが小学校に入学する前年度には、次の4月から通うことになる公立小学校の名前が記載された就学通知書が送られてくる。この就学通知は住民基本台帳から作成された学齢簿に基づいている。小学校や中学校で学ぶ間には、健康診断結果や体力・学力テストの結果、さらには経済的援助の必要な子どもたちに対して支給される就学援助の受給の有無などの業務情報が自治体内に蓄積される。

働いて所得を得るようになれば住民税を自治体に納付するようになり、納税額を決定するための所得情報が自治体に集まる。また、生活保護などの社会保障を受給するためにも、所得情報は必要である。このほか

にも、婚姻・離婚、死亡といった戸籍に関する届出や、転居による住所の転入・転出の届出など、地方自治体はさまざまな行政サービスを提供している。その窓口である区市町村自治体には行政サービスを提供する上で必要な情報が集まり、その実態を記録したものが、自治体における行政データとして日々記録されている。行政データとは、全住民の日々の活動の記録である。

政府統計の総合窓口であるe-Statによると、行政記録情報（行政データ）とは、国の行政機関が保有する各種行政記録情報や地方公共団体が共有する業務記録情報のことを意味する。具体的には、個別の法令の規定に基づいて為される申請や届出、登録や報告等によって国や地方自治体の行政機関に蓄積される情報や、行政機関の日々の業務活動を通じて収集される情報を指す。

地方自治体において、その活動を支える基礎資料が住民基本台帳である。住民基本台帳は、氏名、生年月日、性別、住所などが記載された住民票を編成したもので、住民に関する事務処理の基礎となるものである。住民基本台帳は、国民健康保険・後期高齢者医療・介護保険・国民年金の被保険者の資格の確認や、児童手当の受給資格の確認、学齢簿の作成、生活保護及び予防接種に関する事務といった様々な行政事務に活用されている。この住民基本台帳に紐付けられる形で日々の行政事務の情報が記録されている。

全住民の日々の活動を記録した自治体行政データは、人々の活動を理解し、社会の動きを把握するための情報の宝庫である。近年、自治体の業務を通じて収集される行政記録情報の学術研究への活用が活発になされている。本稿では学術利用という観点から行政データのメリットとデメリットについて述べ、自治体行政データの特徴について、公的統計との違いに着目しながら議論する。その上で、自治体行政データを活用した国内外の学術研究の例をいくつか紹介し、最後に今後の展望を述べる。

2 行政データのメリットとデメリット

行政記録情報として収集されるデータを研究に活用することのメリットは非常に大きいですが、デメリットもある。以下に行政データの学術上の利活用におけるメリットとデメリットをまとめる¹⁾。

(1) メリット

まず、情報の質という観点から行政データのメリットを考えてみよう。行政データはそもそも業務を遂行する上で必要な情報である。そのため、記録される情報には常に高い正確性と客観性が要求される。例えば、国税額や住民税額を確定する際には、所得についての情報が必要であるが、源泉徴収や確定申告を通じて記録される所得額は正確であることが要求される。仮に所得額の虚偽申告をすれば修正申告が要求されたり、追徴課税等の罰則が科されたりする。また、医療保険額の決定に必要な医療行為に関する情報は、専門家としての医師が診療行為についての申告を行っているため、通常はその正確性および客観性は高いと考えられる。万が一診療行為に関して虚偽の申告をすれば、詐欺罪で罰則の対象となる。調査への参加がボランティアである社会調査では、虚偽の回答をしたとしても罰則等を科すことはできず、回答の正確性や客観性が低下する場合もある。正確な回答が要求される行政データは、社会調査と比較して正確性および客観性が高い。

次に情報量の観点からのメリットとして、その悉皆性があげられる。行政データはその業務の対象となるすべての人々についての情報を記録する全数調査である。そのため、抽出調査を主体とする従来の社会調査と比べて、調査の対象となる母集団の全体像を把握することができる。また、調査対象としての母集団が大きな場合には、抽出調査に比べて標本サイズの大きなデータが利用でき、統計的な分析を行う際の検出力が強くなる。特に、標本サイズの大きい行政データには、従来の社会調査における抽出調査では対象になりにくい（例えば、双子や外国人労働者などの）少数派の情報も含まれている可能性が高く、そういった人々を対象とした統計的分析も可能になる場合がある。

情報量の観点からみた別のメリットとして、データの接続可能性の高さがある。行政業務データにはその行政サービスを受ける人の名前やIDが必ず付与されている。この個人を識別する情報が異なる業務間において共通であれば、これらを用いて異なる業務データを紐付けることができる。それぞれの業務はそれぞれ

別の目的を持って遂行されており、そこで収集される業務上のデータは目的により異なるが、それらを紐付けすることができれば、同じ個人に対する情報の種類が増える。行政業務データはその行政サービスを受ける人々に対する全数調査であるため、同じ人が異なる行政サービスを受けていれば、複数の業務データに情報が含まれることになり、異なる業務データの紐付けが可能になる。

接続可能性に関連して、追跡可能性の高さもメリットである。行政データは行政サービスを受けるためには必ず答えなければならない“調査”への回答であり、調査への参加は強制である。そのため、調査への参加がボランティアである社会調査と比較して、無回答やサンプルからの脱落の問題が少ない。例えば、所得税額の決定において所得額の申告は義務であるが、社会調査での所得の質問への回答はボランティアであり、無回答やサンプルから脱落する恐れもある。個人の追跡可能性の高さは、時に世代を超えた追跡をも可能とする。例えば、出生届には子の名前と親の名前の両方が記載されているが、親と子それぞれの行政記録情報を紐付けることによって、親子関係の長期的な分析もできる。

(2) デメリット

研究に利活用するという観点から見て、行政データの持つデメリットもある。まず、各データの包括性の低さがある。そもそも、行政データは業務遂行のために記録されたデータであり、研究のために集められたデータではない。そのため、研究に必要な情報が含まれていないことが多い。例えば、教育と所得の関係を明らかにすることは労働経済学における最も重要な研究課題の一つであるが、所得額の申告データの中には学歴の情報がなく、所得額の申告データのみでは分析できない。一つの業務データにおいて学術上重要な問いを分析することができるケースは稀であり、複数の業務データを紐付けることが分析上不可欠となることが多い。

次に、行政データは翻訳が必要な場合が多い。行政データは行政業務の遂行のための内部資料であるので、行政内の人々に意味がわかれば良い。逆に、研究者をはじめとする行政外の人々にとっては、これらの情報が何を意味するのかがわからないことも多い。例えば、自治体の教育委員会が児童・生徒の意識調査を実施したとしても、その調査項目や調査票が公開されていなければ、行政データに記録されている数字の意味がわからない。そのため、行政データを研究目的で

利活用するためには、それらを研究に用いることができる形に翻訳する必要があり、そのために多くの時間と費用が必要となる。

行政データは包括性が低いために、異なる業務データを紐付けする必要がある。このデータの紐付けのためには、名前や生年月日、生まれた場所といった各個人を識別するための情報が必要であるが、これらの個人情報保護には細心の注意を払う必要がある。北欧諸国のように、各行政データに共通の個人番号が付与されており、それらを識別子として用いることで紐付けが容易に行える場合もあるが、そのようなケースは世界的にも稀である。共通番号がない場合には、性別や誕生日、居住地域（や氏名）といった（標準識別子と呼ばれる）情報で紐付けをしなければならない場合も多い。また、国によってはそのような紐付け自体が禁止されている場合もある。個人レベルでの接続が認められるとしても、個人情報を含む行政データを公開することは認められず、データを分析する際にも個人名の匿名化が必要とされる。さらに、行政データの分析は、セキュリティの高い指定場所において指定された方法によってのみ認められる場合が多く、「データと踊る」、つまり、自由にデータを触りながらさまざまな仮説を考えつつ分析を進めることが難しい場合も多い。このように、行政データを研究に利活用するためには、個人情報保護のための法制度の整備から情報セキュリティシステムの構築まで、さまざまな追加的費用が必要となることも多い。

(3) 公的統計との比較

国の行政機関や地方公共団体は、行政利用のためだけでなく、社会全体で利用される情報基盤としての公的統計を作成している。この公的統計を作成するために、さまざまな統計調査が行われており、2009年4月に施行された新・統計法のもとで、これらの公的統計調査は学術目的での二次利用に活用されている。これらの公的統計調査と比較して、行政データ、特に自治体の持つ行政データのメリットは何であろうか。

国の行政機関が行う調査は基幹統計調査、一般統計調査、及び、業務統計調査に分けられる。その中でも、『国勢調査』をはじめとする基幹統計調査は正確な統計を作成する必要性が特に高く、基幹統計調査の報告（回答）を求められた者が、報告を拒んだり虚偽の報告をしたりすることを禁止する報告義務（統計法第13条）、及び、罰則規定（統計法第61条第1号）が定められている。これにより、記録される情報の正確性と客観性が担保されている。また、『国勢調査』

のような悉皆調査であれば、悉皆性も確保される²⁾。自治体の行政データと比べて、国の行政機関が行う調査は対象が日本全体と大きく、大きな標本サイズを駆使した統計分析ができるという意味では各自治体の行政データよりも優れている。ただし、地方公共団体のデータベースのネットワーク化やオープンデータ化の促進によって複数自治体の行政データを一括して分析することができれば、克服できる点であるとも言える。

国の行政機関が行う調査のうち、基幹統計調査以外の一般統計調査は報告の義務が課されていない。そのために、行政データに比べてデータの悉皆性に欠け、場合によっては正確性や客観性も注意して分析する必要がある場合もあるかもしれない。この点は地方公共団体が実施する統計調査も同様であり、これらの一般統計調査は社会調査と同様の問題を持ちうると言える。

次に、国の行政データと自治体行政データを比較してみよう。もし国の行政データと自治体の行政データが同じ情報を含んでいれば、標本サイズという観点からは国の行政データの利用が圧倒的に有利である。しかしながら、データによっては自治体のみが保有し、国の行政機関は保有していないものもある。例えば、国税庁は確定申告を行った人々の所得データは保有しているが、源泉徴収のみで年末調整を受けている人々の所得情報の詳細は自治体が保有しており、住民の所得情報を包括的に保有しているのは自治体となる。

また、行政データの学術利用においては、異なるデータの紐付けが重要であるが、自治体内であれば、標準識別子を用いた横方向の紐付けが比較的容易である。公的統計調査では、名前などの標準識別子は統計調査票には通常含まれておらず、異なる統計調査における個人レベルの情報を別の統計調査に接続することが困難な場合が多い。たとえ同じ統計調査内でも、識別子が限られており接続は困難で、そもそも個人の特定は個人情報保護のために禁止されているので、識別子はマスクされる傾向が強い。

最後に、縦割り行政の弊害の違いがある。異なる目的の調査やデータを紐付けるには、異なる省庁や部署の壁を越える必要がある。例えば、教育分野と福祉分野のデータを紐付けるためには、国レベルであれば文部科学省と厚生労働省の間の壁を越えなければならないが、自治体の場合には教育課と福祉課の間の壁であり、省庁間に比べると壁は低く、異分野間のデータの紐付け可能性は自治体の方が高い。

以上、自治体行政データの学術利用上のメリットとデ

メリットについて、社会調査や公的統計調査との比較において議論した。次節では、行政データを活用した国内外の研究例を紹介する。

3 自治体行政データを用いた分析の例

行政データを活用することで、様々な仮説を検証する研究が、ハーバード大学のチェッティ教授らのチームによって行われている。Chetty, Friedman and Rockoff (2014) では、アメリカのある自治体の教育行政データを、税の申告書のデータと紐付けることによって、小学校の時にどのような先生に教わったかによって、その子どもたちが大人になった時にどれだけの所得を得ているのかを調べている。その結果、学力を伸ばす効果が下位5%先生を平均的な先生と交代させると、その後の生涯収入がクラス全体で25万ドル多くなることを明らかにしている。この他にも、国税データを用いて、世代間の所得の相関（世代間所得移動度）を調べた研究（Chetty et al. 2014）や、幼少期を過ごした地域が所得移動度に与える影響を調べた研究（Chetty and Hendren 2018）など、行政データをさまざまな情報と紐付けしながら、学術上重要な研究を数多く行っている。

日本においても行政データを活用する動きがあり、自治体独自の教育行政データベースを構築し、それを用いて様々な研究が行われている。その例として、学級児童数（クラスサイズ）が学力と学級閉鎖に与える影響を調べた二つの研究を紹介する。田中（2020）は匿名の自治体の教育行政データを用いて、クラスサイズの縮小が児童の学力に与える影響を調べ、クラスサイズの縮小は小学校児童の学力、特に算数の学力の伸びに対して小さいながらも統計的に意味のある正の効果を持つことが明らかにしている。早稲田大学の及川雅斗助教授の研究（Oikawa et al. forthcoming）はクラスサイズとインフルエンザによる学級閉鎖の関係を調べ、クラスサイズの縮小は児童生徒間の「ソーシャルディスタンス」の改善を通じてインフルエンザによる学級閉鎖を減らすのに一定の効果を持つことを明らかにしている。これらの日本のデータを用いた研究は、海外の研究のように子どもたちの将来所得との関係の分析までには及んでいないが、行政データの蓄積を継続し、紐付けできれば、さまざまな分析が可能になるであろう。

4 自治体行政データ活用の可能性と限界

自治体行政データは、従来の社会調査や政府統計調

査とは異なった性質を持ち、学術研究を進めてゆく上で大きなメリットをもたらす。しかし、それを現実のものとするためには、克服しなければならない課題が残されている。自治体行政データを活用することによって、学術研究の促進を図り、より良い社会制度の設計を目指すことは、急速な人口減少に直面する日本において最重要課題の一つである。すでにある行政データを利活用するのみならず、将来の分析に備えてそれらのデータを整備・蓄積し、活用する制度を今から作っておくことは、将来世代に対する我々の責務である。

- 1) ここでの行政データのメリットとデメリットに関する議論は、フィグリオらによるまとめ（Figlio, Karbownik and Salvanes 2016）と別所ほか（2019）を参考とした田中（2021）を大幅に改定したものである。教育行政データの研究利用例およびデータベースの構築についても田中（2021）を参照のこと。
- 2) しかしながら、悉皆調査の回収率は必ずしも100%には届かず、令和2年の「国勢調査」の回収率は約8割であった。

参考文献

- 田中隆一（2020）「根拠を活用した教育政策形成へ向けて——自治体教育データを用いたクラスサイズ縮小効果の検証」『社会保障研究』第5巻、第3号、pp. 325-340。
- （2021）「行政データで明らかにする教育の効果」『経済セミナー』No. 719、4-5月号。
- 別所俊一郎・野口晴子・田中隆一・牛島光一・川村顕（2019）「子どもについての行政データベースの構築」『フィナンシャル・レビュー』第141号、pp. 106-119。
- Chetty, R., J. N. Friedman and J. E. Rockoff (2014) “Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood,” *American Economic Review*, Vol. 104, No. 9, pp. 2633-2679.
- Chetty, R., N. Hendren, P. Kline and E. Saez (2014) “Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States,” *Quarterly Journal of Economics*, Vol. 129, No. 4, pp. 1553-1623.
- Chetty, R. and N. Hendren (2018) “The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects,” *Quarterly Journal of Economics*, Vol. 133, No. 3, pp. 1107-1162.
- Figlio, D., Karbownik, K. and Salvanes, K. G. (2016) “Education Research and Administrative Data,” *Handbook of the Economics of Education*, Vol. 5, pp. 75-138.
- Oikawa, M., R. Tanaka, S. Bessho and H. Noguchi (forthcoming) “Do Class Size Reductions Protect Students from Infectious Disease? Lessons for Covid-19 Policy from Flu Epidemic in Tokyo Metropolitan Area,” *American Journal of Health Economics*.

たなか・りゅういち 東京大学社会科学研究所教授。主に「計量経済学の第一歩——実証分析のススメ」（有斐閣、2015年）。教育経済学、労働経済学専攻。