

行政データの活用



早稲田大学教授 野口 晴子

1. 行政データとは

行政データとは、様々な行政目的のために、国や地方自治体による行政活動を通して収集されるデータである。たとえば戸籍とは、「日本国民の国籍とその親族の身分関係を戸籍簿に登録し、これを公証する」システムであり、市区町村、あるいは、海外居住者については日本大使館や領事館によって、当該情報の処理や管理がなされている行政データである。

また、私たちは、現住所地である市区町村に対して、子どもが生まれたら出生届、結婚や離婚した際には婚姻届や離婚届、

親族が死亡すれば死亡届、引越時に際しては転入届を提出し、そうした記録が住民基本台帳に反映され蓄積される。住民基本台帳は、国籍とは無関係に、外国人も含め日本に暮らす人々の居住関係を公証する制度であると同時に、選挙人名簿への登録、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、児童手当の受給資格の確認、学齢簿の作成、生活保護・予防接種・印鑑登録に関する各種事務処理に利用されている。

その他にも、私たちが、病院や診療所、介護施設や介護サービス事業所から、医療や介護の

提供を受けた際には、サービス提供者が保険者に提出する報酬明細書(レセプト)が、また、罪を犯せば、警視庁や警察庁の犯罪データベースに記録として残され、蓄積される。

つまり、行政データは、シカゴ大学のジョン・リストがいうところの自然発生的な観察データとして、特定の行政目的の対象となる人々の意思決定が全数で集積されているデータであり、実験室という閉鎖空間において、限られた人やモノを対象として行うラボ実験の究極に位置付けられる情報群であると考えることができる。

2. 行政データの長所と短所

政策研究であるかにかかわらず、研究者が行政データに着目するいくつかの理由がある。第一に、特定の行政目的に限定されているものの原則、当該目的の対象者の全数についてデータが収集されるため、標本抽出の過程で発生する選択による偏りを回避することができ

る。第二に、たとえば窓口業務での入力ミスや申請者による記載ミスなど事務処理上の過誤を除けば、研究者による独自調査と比べ、回答者自身による主観が入り込む余地が少なく、回答による偏りが小さいと考えられる。第三に、既にルーティンとなっている行政活動の一環としてデータが収集されるため、膨大な初期投資がかからない。そして最後に、情報収集の主体である国家や行政機関、あるいは第三者によるデータの目的外の二次利用について、倫理的問題が伴うことは必須であるが、少なくともデータ収集にあたっては、倫理上の問題がない。

もっとも、「倫理上の問題がない」と断言するには大いに留意を要する必要があるかもしれない。なぜならば、歴史上、個人や企業などの経済主体による自由な意思決定を阻害し抑圧することを目的に、国家が情報を収集し活用した事例には事欠かないからである。だからこそ、情報収集の目的や活用に対し、相応のモニタリングコストをか

ける必要がある。

他方、行政データには短所もある。特定の行政データから得られる情報は極めて限定的であるという点である。その一例が「レセプト情報・特定健診等情報データベース(NDB)」である。NDBには検査・治療内容・診療報酬点数が網羅されているが、受診者属性については、最も提供要件のハードルが高い「特別抽出」であっても、性・年齢・居住市区町村程度の情報しかなく、たとえば、家族構成や所得など社会経済的な状況に関する情報が含まれていない。これについては、介護に対するレセプト情報も同様である。

また、行政データには、直接住民の利害に影響する極めてプライベートな情報が含まれているため、そうした情報が個別ベースで研究者に提供される際には、個人情報保護に関する倫理規定が課されることになる。研究者としては、自身の研究目的を達成するため、個人情報とできるだけ幅広く取得したいと考えるのが当然であるが、情報

量が多ければ多いほど個人が識別されるリスクが高まるため、必然的に、研究者に課される倫理上の責任は重くなる。

日本では、情報が漏えいした場合、情報の提供を受ける研究者よりも、国や地方自治体など情報を提供する側に対する法的・社会的制裁の方が大きい制度設計になっていることから、情報提供者にあまり多くの情報を提供したくないというインセンティブが働く可能性がある。

こうした行政データの短所や、二次利用にあたっての倫理規定については、日本だけの話ではなく、日本よりも行政データの利活用が盛んに行われている欧米にも共通する普遍的な課題である。

3. 米国の経験に学ぶ

情報収集の目的が異なっているため、単体では限られた情報しか得られないという行政データの短所については、複数の行政情報を照合させることで、研究の遂行に必要な情報を追加的

に得るという手段が考えられる。複数の行政データを照合させるためにどうしても必要不可欠な変数が、個人に割り振られた「共通番号」である。

米国では、こうした共通番号として、社会保障番号(SSN)が用いられていたが、1996年に、「医療保険の携行性と責任に関する法律(HIPPA)」が制定され、米国民健康社会福祉省は、HIPPAに基づき、健康情報に関するプライバシー及びセキュリティを策定した。第一に、匿名データを提供される側が個人を再度識別するリスクが最小化され、専門家が統計的・科学的手法により確認可能なこと、第二に、氏名、詳細な住所情報、生年月日、電話・FAX番号、メールアドレス、SSN、医療レセプト番号、健康保険番号など、個人を識別可能な情報を削除した後、残り情報を個人が識別できないことを確認可能なこと、というルールである。HIPPAが打ち出したこうした方向性、とりわけ、匿名化の方法をめぐっ

て、研究者の間で賛否両論、様々な議論が医学系の専門誌上で現れも続けられている。

しかし、プライバシールールは、個人の識別が完全に不可能であることを保障するものではない。たとえば、全米保健医療統計センターなど政府系の研究機関では、共通番号を用いずに行政データを照合し、照合後の匿名化されたデータ上で再度個人が識別されることのないようにする手法の開発、個別の行政情報を照合するにあたっての法的妥当性の検討、そして、様々な研究目的に合わせた照合データの研究者への提供など、HIPPAのプライバシールールに沿った形で研究支援体制の構築が行われている。

日本でも、国民と法人に対しマイナンバーが割り振られたが、それを行政の現場や研究においていかに活用するかについては、今後の大きな課題であり、過去約20年間にわたる情報の有効活用と個人情報保護に関する米国の議論に学ぶべきことは多い。