

社会動向レポート

EBPM 推進にむけた、データ整備に必要な「5つの視点」

—北欧諸国・台湾及び国内2事例における行政記録情報等の活用の観点から

社会政策コンサルティング部

主任コンサルタント 森安 亮介

コンサルタント 利川 隆誠

政策の質を高めるために、行政や公的機関等の持つ統計・行政記録情報を活用する EBPM (Evidence Based Policy Making) に多くの関心が寄せられている。ただし、他国や他自治体を模倣しただけではうまく機能しない。行政機関等が行政記録情報のデータ整備・データベース構築を検討する際には、データ収集の持続性やデータ利活用の継続性を見据えた「5つの視点」が求められる。

1. はじめに

「データの世紀」と言われる21世紀も四半世紀を経ようとしている。データに基づく経営、データドリブンマーケティングなど、データの利活用は人々の意思決定や組織の提供価値向上において有用な手段として浸透しつつある。そしてそれは政策領域も例外ではない。

1990年代にイギリスやアメリカで進められたエビデンスに基づく政策立案(Evidence Based Policy Making。以下 EBPM という。)は、いまや先進諸国に広まっており、日本においても経済財政運営と改革の基本方針(骨太の方針 2017)を基軸とし、2018年頃から全省庁で推進されている。EBPM の浸透によって、これまで以上に適切な政策の選択・遂行が期待される上、データの拡充が進めば、これまで発見しにくかった社会課題を見出し、より効果的かつ効率的な打ち手に結びつけることも可能となる。例えば森安(2022)が例示したように、医療・教育・雇用

など従来の政策領域をまたいだデータを組み合わせることによって、一見医療政策上の問題に見える事象の真因を教育に見いだすような発見にもつながる。しかし現状、そうしたデータが十分に整備されているとは言い難く、まさに政府や各自治体でデータの在り方が議論される黎明期にあると言える。

筆者らは2019年以降、厚生労働省や文部科学省など中央省庁向けの EBPM 推進支援に携わってきた。具体的には、各種エビデンス収集支援や政策評価分析、ロジックモデル作成支援、各種研修、マニュアル作成を行ってきた。また、幾つかの自治体にも研修・勉強会等を含めた各種側面支援をしてきた。そこで直面する大きな課題が統計・データの整備である。仮に行政職員が EBPM をよく理解し、意欲的に取り組もうとしても、その土台となるデータの不足によって、取組みが制限されるケースを数多く目にしてきた。

では EBPM の土台となる統計・データの構

築に向けて、何が必要なのだろうか。本稿では、筆者らが取り組んできた行政向け EBPM 支援の経験に加え、2021年度総務省統計委員会担当室事業において行った国内外の事例調査「公的統計の国際比較可能性に関する調査研究(社会統計編)」(総務省 2022)⁽¹⁾で得た知見も織り交ぜて検討する。

2. 先進諸国にみる、統計・データ整備の背景

EBPM は政策目的を明確にした上でエビデンスに基づく政策を立案する。その過程ではデータをもとに政策課題の特定や政策効果の検証を行う。このとき必要となるデータが、同一対象の変化を経年的にとらえたデータ(パネルデータ)である。なぜなら、一時点の断面的なデータでは、政策とその効果の因果関係にまで迫る検証が非常に困難なためである。因果関係を統計的に推定するためには、政策対象である個人や家計、学校といった層の政策前後の変化を、政策非対象の層と比較することが望ましく、精度の高い効果検証に必要不可欠である。中でも、行政機関や公的機関が業務上すでに取得している情報(行政記録情報)を用いることで、時間的・金銭的なコストをおさえてデータを得ることができる(森安 2022)。実際、例えば北欧諸国に代表されるような先進諸外国では、行政記録情報を国民番号等で紐づけた行政記録情報を活用したデータベースが構築され、いくつもの分析が行われている⁽²⁾。

諸外国の先進事例は、模範的なデータ像として、しばしばデータの構造や技術が注目される。しかし我が国や各自治体への応用を考える際、そうしたデータの中身以上に大切なのは、運用やルール、関係者の理解・インセンティブなどを歴史的・社会的・制度的な背景とともに捉えることにある。後述のように、北欧諸国におい

ても長い年月をかけて成熟し、人々の日常生活にも浸透しているからこそ持続的に機能している側面がある。こうした問題意識をもとに、以下2つの事例を通し、データの整備や利活用の背景に目を向けていく。

(1) 北欧諸国における、行政記録情報を用いたデータベース構築とその背景

行政記録情報を用いたデータベース構築の模範事例と目されるのが北欧諸国である。例えば、当社コラムでも紹介したように、デンマークでは国内在住者を管理する CPR 番号(Central Persons Registration)を起点としたデータベースが構築されている(栗山 2022)。また、教育先進国でも知られるフィンランドでは、教育文化省が全国約300自治体の教育機関から生徒に関する情報を業務データとして収集し、統計局が個人 ID によって管理している。

しかし、こうしたデータベース構築の背景には、長い歴史の積み重ねや社会的・文化的な背景がある。北欧諸国では16世紀ごろから教会が住民台帳を作成しており、住民の属性情報が収集・把握されてきた⁽³⁾。また、個々人に付与される固有番号(個人識別番号)についても百年近い歴史を有しており、市民生活にも浸透している。例えばデンマークでは、個人識別番号の整備は1924年から開始され、1968年に前述の CPR 番号制度導入につながっている。以降、医療情報との紐づけ(1977年 国民患者医療記録システム導入)、2007年には行政機関・個人・法人間での円滑な情報交換に向けた電子署名制度などの変遷を経ている。CPR 番号は、医療・福祉・その他市民生活を営む上での行政サービスや銀行口座開設・融資・賃貸契約などの場面でも市民に活用されており、人々の生活に根付いた存在になっている。

さらに、個人情報に対する考え方も日本のそ

れとは大きく異なる。例えば今井(2017)によると、北欧諸国では、情報については透明性を優先するという文化的な国民のコンセンサスが存在するという。例えばスウェーデンの国税庁では、住民情報登録に名前・住民登録番号・住所・出生地・スウェーデン国籍の有無・配偶者や子どもの情報・死亡その他の理由による住民登録からの離脱状況・合計収入等の情報が蓄積されているが、このうち住所・氏名・住民登録番号等は公開可能で、企業販売までされている⁽⁴⁾。

このような、個人識別番号の市民生活への浸透や、個人情報に対する一定のコンセンサスは、データ連携の社会的合意を得るための調整コストが低いことを意味する。もちろん近年の個人情報保護への関心や EU の法的な影響から、今後変化する可能性もある。しかし、個人情報への意識が高まる以前の段階で、すでに市民生活に個人識別番号が浸透していた点は見逃せない時系列的事実であろう。なにより、半世紀以上もの時間をかけて、まず市民の利便性を高める仕組みが構築され、市民生活にも浸透した後に、結果としてデジタル化によるデータ活用につながっていった。こうした前後関係は、データの構築や持続的な運用を考える上で、示唆深い事実である。

なお、行政記録情報を活用したデータベース構築・運用のコスト負担と国民の納得感を考える際、北欧諸国が高福祉国家であることも重要な要因だと考えられる。福祉国家として適切な行政サービスを提供するためには、国民一人一人の就業状態や生活・家族の状態などの基礎情報が必要になるためである。これら情報の行政による把握は国民生活向上にも直結することから、データベース構築への投資が合理的であり、そのベネフィットも大きいこととなる。

(2)台湾の医療データベース構築とその背景

では、北欧以外のケースではどうだろうか。データを活用したデジタル民主主義を標ぼうする台湾の事例から、2002年運用開始の医療データベース NHIRD (National Health Insurance Research Database) を参考にしたい。

台湾では北欧諸国と同様に在住する者はすべて全民健康保険証という国民証(国民身分証統一番号)と共通した10桁の番号が記載された IC カードを有している。この IC カードは患者の医療情報が記録され、診察・検査・治療内容・投薬などの情報が衛生福利部(日本の厚生労働省に相当)の NHIRD に集約される。NHIRD を活用し、政策のための各種分析のほか、匿名化した上で台湾の研究者にも研究目的で提供されている。さらに、政府のプロジェクトであれば、NHIRD は国民身分証統一番号を起点として他分野のデータベースとの連結も可能である。台湾では国民身分証統一番号を有する社会経済的(所得、加入保険等)、文化的行動因子(結婚の有無、家族構成等)を格納したデータベースなども構築されている⁽⁵⁾。これら複数のデータベース連結によって、社会経済状況や文化的行動など様々な観点からの分析が可能になっている。

以上のように行政記録情報等を活用した豊富なデータベースが構築されている台湾ではあるが、しかし、運用面等とあわせて捉えると別の側面も見えてくる。筆者らは前述の総務省(2022)にて、台湾の同データベースを中心に現地有識者にヒアリングを行った。結果、見逃してはならない以下2つの点も明らかになった。

第一に、これらの医療データは機微な情報を含んでおり、利用範囲やプライバシー保護等について社会的合意を完全に得られているとはいえない点である。実際、NHIRD のデータ活用等におけるプライバシーへの配慮は市民団体から複数回の訴訟や反発をまねいている⁽⁶⁾。

第二に、そうした影響もあってか NHIRD を研究者等が利用する際、厳重な管理・運用がなされている点である。例えばデータ利用を希望する場合、研究計画を示した申請書を提出し、承認されることが必要となる。そして利用時には、データ利用料が発生する。一例ではあるが、過去5年のがん患者データを使用した場合、データ抽出費に10,000米ドル、加えて一日のデータ利用セッションタイム費が加算されるという。また、データ分析はデータセンターなど特定の数か所でのみ作業可能となっている。センターの出入りは厳重で、研究者が自身のペンや紙を持ち込むこともできないという。分析結果は全て点検され、個々のデータを外に持ち出すことはできない。センター退出時には、作成したメモ等をセンターに預け、次回入室時に当該メモを受け取るような運営となっている。

このように、NHIRD はデータの内容は豊富であるものの、データ利用時に制限をかけている様子が伺える。台湾の一部研究者から改善希望の声も上がっているものの、市民の受容性に鑑みると、データの内容とのトレードオフになっている関係性が示唆される。

3. データ構築を考える「5つの視点」

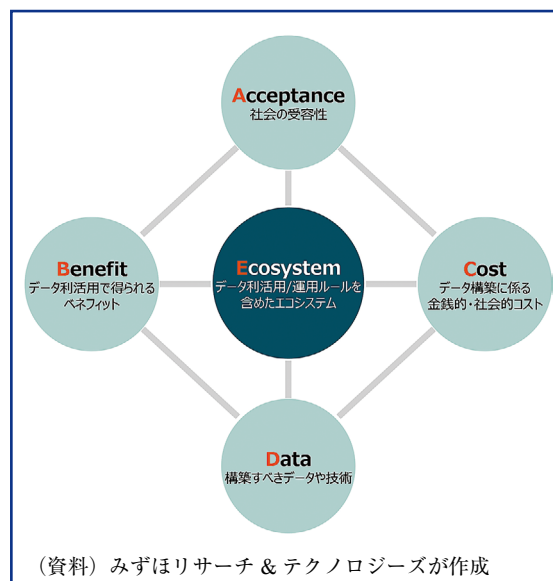
以上、北欧諸国や台湾の事例を確認した。北欧事例からは、個人番号に関する歴史的背景に加え、個人識別番号の浸透、個人情報への寛容さ、さらに福祉国家である点など社会的・制度的背景があるからこそ、行政記録情報のデータベース化や利活用が可能になっていた。また台湾事例では、データの豊富さと引き換えに、高額な利用費負担や厳格な運用制限をかけている様子も伺えた。その背景に社会的な合意による影響もあることが示唆された。こうした事例を踏まえると、統計・データの在り方は、歴史的・社会的・制度的背景をふまえた社会の受容性の

影響を色濃く受けながら、データ利活用によるベネフィットとデータ整備・運営コストの妥当性、そして利活用の運用・管理方法などと密接に関係しながら規定されるものと考えられる。

これら EBPM に必要なデータ整備についてこれまで受けてきた相談等もふまえ、筆者らはデータ構築には以下「5つの視点」で検討すべきだと考えている。

ここで強調したいのは、社会の受容性 (Acceptance) を土台に、「整備すべきデータや技術 (Data)」だけで検討するのではなく、ベネフィット (Benefit)・コスト (Cost) などとの関係性の中で検討することである。そして、関連諸制度やステークホルダーの相互関係もふまえた運用とともに、いわばデータを取りまくエコシステム (Ecosystem) として捉えることが必要である。以下、それぞれの視点について、総務省 (2022) で示された国内事例も踏まえ、詳述したい。

図表 1 サステナブルなデータ利活用に向けた「5つの視点」



(1) Acceptance：社会の受容性を土台に考える

行政記録情報を活用したデータベース構築でまず前提になるのは、データに関する関係者の受容性である。データベース構築やデータの利活用は、個人のプライバシーや人権の配慮、セキュリティ面のみならず、倫理面や市民感情、世論、歴史的経緯や社会的背景などとも密接に関連する。そのため市民やステークホルダーと一緒にした議論が必要不可欠である。我が国では、令和4年「個人情報等の適正な取扱いに関する政策の基本原則」（個人情報保護委員会）等で個人情報保護と有用性への配慮、データガバナンス等については指針が示されているものの、個別分野の運用や責任の在り方などはそれぞれに検討する必要がある。なぜなら、それらは分野ごとに既存制度との関連性や諸慣行の積み重ねによって形成されるためである。データに関わる市民が、データ利活用範囲やリスクなども含めてどこまで理解して許容しているのか。その許容範囲に応じた検討が必要である。もしも誤解や情報不足などによって受容性に課題がある場合には、周知徹底やコミュニケーションを行い、丁寧な合意形成を行うことが求められる。

そのような合意形成が求められる中、日本で行政記録情報を活用したデータベースの構築事例として知られるのが大阪府箕面市の「子ども成長見守りシステム」である。このシステムも、丁寧な対話を通し、市民やステークホルダーから信頼を得ながらデータの利活用がなされていった経緯が確認できる。

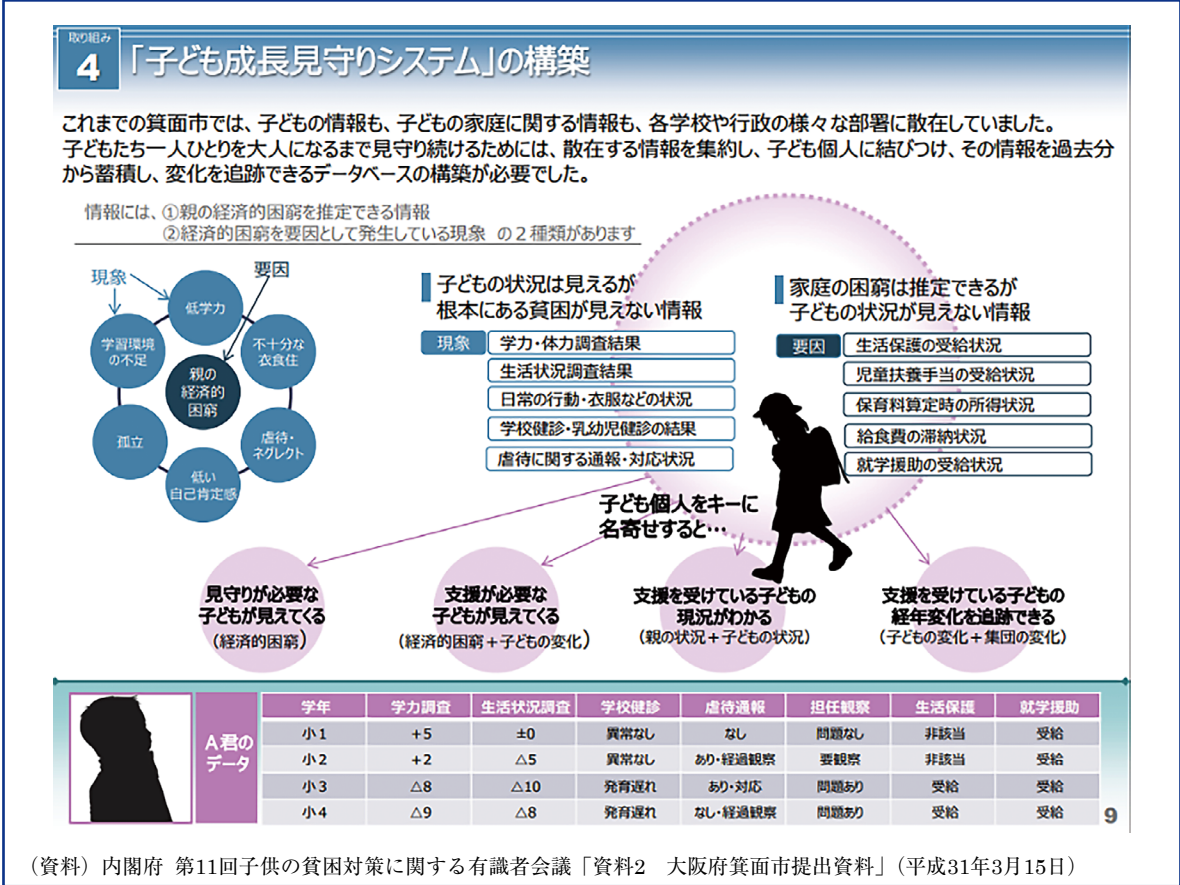
同システムは、教育分野の情報(学力調査・体力調査・生活状況調査・学校検診・登園状況など)と福祉分野の情報(生活保護・就学援助・給食費の滞納状況・虐待に関する通報状況など)を融合したシステムである。教育と福祉の情報が

個々の生徒ごとに紐づけられているため、例えば学力低下懸念の生徒の早期発見など、成長の側面支援に活用されている⁽⁷⁾。着目すべきは、データベース構築の過程である。筆者らは前述の総務省(2022)にて、同システム構築を牽引した当時の市長 倉田哲郎氏に話を伺っている。その結果、対話を通した次の2つの信頼獲得が明らかになった。

第一に学校関係者との合意である。2016年に箕面市に新設された「子ども成長見守り室」が、データへの理解や信頼獲得に大きく寄与したと倉田氏はいう。同室は、福祉部門と教育部門に分かれていた子ども関連部門が統合され、いわば子ども貧困対策のコントロールタワーを担っている組織である。システム構築当初、同室担当者が各学校に足しげく通い、学校長を含む学校関係者と密にコミュニケーションを取っていた。学校現場の持つ勘と経験がデータで補完されることで、今まで気づけなかったリスクが明らかとなった。その結果、当システムは徐々に信頼され、学校教育の中にも取り入れられていったという。学校にシステムが普及した現在も、密なコミュニケーションは続いており、市内の全20の小中学校それぞれと年2回ずつ会議を行っているほか、年10回程度関係課室やセンターなどと情報共有会などを実施しているという。

第二にデータの取り扱いに関する合意である。システム構築時、専門家も含む審査会の諮問を経て、個人情報保護条例の改正が行われている。倉田氏によると、改正前の個人情報保護条例であっても同システムへの情報活用は可能であった。しかし改めて審査会の場を設け、市民や関係者等との対話を行い検討の結果、既存の条例でも問題はないものの、行政記録情報を活用したデータベース構築後の運用を見据えた条項の追加等が望ましいとの結論を得て、住民合意のため市議会を通じた条例改正まで行っている。

図表2 子ども成長見守りシステムのイメージ



これらの取り組みによって、行政現場と住民双方が安心して運用できることにつながったと倉田氏は語る。

(2) Benefit・Cost: ベネフィット・コストを吟味した上で、データの在り方を検討する

まず市民や関係者の受容(Acceptance)を紹介したが、次に検討が必要なのがベネフィットとコストを吟味したデータベース構築や運用である。ここでも国内事例が存在する。日本を代表する厚生労働省のレセプト情報・特定健診等情報データベース(以下、「NDB」という。)の事例である。

NDBは特定健診・保健指導の結果と医療処置を行った際に請求されるレセプトデータの2

種類の行政記録情報を格納したデータベースである。本データベースは2005年度の「生活習慣病健診・保健指導の在り方に関する検討会」での議論を基礎に、生活習慣病の予防・改善のための特定健診・保健指導の効果検証を目的に構築された。当時の診療記録は各医療機関、保険者ごとに管理されていたが、判定基準や評価項目の不統一、個人追跡の未実施など、施策の適切な効果検証が困難であった。そこで、特定健診・特定保健だけでなく、保険者から提出された全ての医療レセプトデータを蓄えるデータベースとなったのである。

NDBのあり方には情報技術や公衆衛生等の各分野の有識者が集い検討された。この検討過程で示唆深いのは、ベネフィット・コストを踏

まえて必要なデータや運用フローが形作られていった点である。例えば、2007年7月から設置された「医療サービスの質向上のためのレセプト情報等の活用に関する検討会」では、レセプトデータ、特定健診・特定保健指導データの収集方法等が議論された。その報告書(2008年2月)には以下のような記載がある(図表3)。

上記報告書では、NDB構築の目的として、正確なエビデンスに基づいた医療サービスの質の向上等を図ることが明記されている。目的が明確になることで、収集すべき情報項目も明らかになり優先度も決まる。さらに、収集すべき情報が明確になることで、管理・運用の在り方も規定される。例えば病名に関する情報項目は、種類によっては非常に機微な情報ではあるものの、医療施策や医療費の判断ができるようになることは政策的に非常に大きなベネフィットである。しかし、例えば国内に数名しかいない希少疾患を抱える患者の情報が病名を通じて他者に閲覧可能となってしまうことは、個人情報の流出や悪用につながるリスクにもなり得る。言

うまでもなく、倫理的な側面でも問題がある。そのような事態を避けるために、例えば厳格な管理を徹底する環境構築のコストはかける必要がある。ただ、その線引きをどこに置くかといった、情報収集におけるベネフィットとコストのバランスやトレードオフに関する運用・管理方法の可能性については継続して検討されている⁽⁸⁾。

ベネフィット・コストを踏まえた検討によるデータベース構築の目的精緻化は、一見すると当たり前のように思われるかもしれない。しかし現実的には、行政機関等で既に保有する情報を前提に、その延長上でデータベース構築が考えられがちである。重要なのはバックキャスト的に考えることである。すなわち、利活用の目的を明瞭にし「誰にどのようなベネフィットがあるか？」を明確化することによって、管理・運用方法も含めたあるべき姿から逆算して考えることにある。そうすることで、必要なコストや対処すべきリスクの優先度も初めて鮮明になってくるためである。

図表3 NDB構築時のレセプトデータ等収集・分析で議論された主な論点

3 レセプトデータ等の収集・分析に当たっての主な論点

(4) 患者等の個人情報は削除するものの、医療費適正化計画の作成等のために分析上必要な情報として病名等慎重に取り扱うべき情報が含まれているデータを国が収集するに当たっては、**収集・分析によるメリットと収集されることによるデメリットを比較した場合に、メリットが上回っている必要がある。**

しかるに、上記のようにすべてのレセプトデータ及び特定健診等データを収集することにより、次のような分析も含めた活用が可能となり、正確なエビデンスに基づいた効果的・効率的な施策を実施し、医療サービスの質の向上等を図ることができると考えられる。

- ① すべてのレセプトデータを用いることにより、**詳細な分析が可能となり、医療費の実態を詳細かつ正確に把握することができる。**
- ② また、**同一人物を同定したうえで、特定健診等データを経年的に分析することにより、生活習慣病対策による生活習慣病の発症・重症化の防止効果等を評価することができる。**
- ③ さらに、**レセプトデータ及び特定健診等データを突合することにより、生活習慣病対策が医療費に及ぼす影響等について評価することができる。**

(資料) 厚生労働省 「医療サービスの質向上のためのレセプト情報等の活用に関する検討会」 報告書(平成20年2月7日)。
なお、太字等の強調は筆者らの編集によるもの。

(3) Data・Ecosystem：データの収集・管理運営・活用など関係者の循環を捉え、エコシステムとして検討する

「5つの視点」のうち重要なのは、データ単体だけで考えないことである。データ整備に付随するコストやリスクについては、運用やルール、関係者間との相互関係なども含めて包括的にとらえることが肝要である。例えば前述の台湾では、機微情報をデータに含めるかわりに、研究者の利用ルールに厳格な制約条件を課していた。また、北欧諸国においてもデータ利用に関する不手際や不正があった場合、利用した研究者等が法的な罰則を受ける仕組みになっている。こうした事例からは、関係者との相互関係や運用を通じてうまく互惠関係を構築していることが伺える。データの作り手側・使い手側・行政等機関・そして市民といった、各ステークホルダーについて、個々の持続性はもちろん、法的側面や運用方法なども含めた相互関係から補完し合うような、いわばエコシステムとして一体的に捉える視座が求められる。データだけを模倣してもうまく機能しないのはこのためである。

実際、国際的にみても、北欧諸国のような個人識別番号に基づくデータベース構築以外の方法も存在する⁽⁹⁾。例えばオーストラリアである。詳細は総務省(2022)を参照いただきたいが、個人識別番号制度自体が存在していないオーストラリアでは、医療・社会保険・税の3つのデータセットを基礎データとし、匿名化された名前・住所などの個人属性情報をマッチングキーとしてデータが接続されている。さらに、健康・教育・所得や税・雇用・人口統計(センサスを含む)の情報を時系列で組み合わせたデータのリンク(Multi-Agency Data Integration Project; MADIP)が進められており、将来的にはMADIPを、法人に関する税・貿易・知的財産を法人番号によってつなぎ合わせたデータ(Business

Longitudinal Analysis Data Environment)ともリンクさせる方針という。このように、国の統計機構や制度に即して行政記録情報を活用したデータベース構築や運用方法が形作られる。

4. 社会環境や分野・制度・ステークホルダーに即したデータベース構築の検討を

以上、国内外の事例も参照しながら「5つの視点」を紹介した。各事例から浮き彫りになるのは社会の受容性やコスト・ベネフィットなどに関連しあいながら、あるべきデータやその運用の輪郭が形作られるといった過程であった。

最後に、行政記録情報を活用したデータベース構築時にすべての要素を規定することが困難なケースにおいて最低限必要な対応について補記したい。行政担当者によっては、「5つの視点」全ての検討・実施が困難だったり、より早い時間軸での対応に迫られているケースもあったりするためである。筆者らは拙速な推進は避けるべきだとは考えているものの、来る将来に備えておく観点で推奨しているのが「マッチングキー」の対応である。マッチングキーとは、異なるデータ同士をつなぎ合わせる際の鍵となるような情報項目で、異なる統計データで共通の項目を持つような項目を指す。例えば先述のNDBでは「保険者番号・名前・生年月日等」を組み合わせたIDをマッチングキーとしていた。実はマッチングキーを用いた接合の障壁となるのは、法的・技術的な側面もさることながら、運用面の課題が大きい。例えば、記入情報の表記揺れや個人情報の変更などである。技術的には、暗号化⁽¹⁰⁾や匿名化処理によって個人情報の特定を防げるが⁽¹¹⁾、この時平仮名・カタカナ・漢字・数字・漢数字などの表記が異なるとマッチングキーとして機能しない。匿名化処理によって代替的に付与される個人IDについて、表記ゆれや情報変化によって別IDの生成を招

図表4 サステナブルなデータ利活用に向けた「5つの視点」に照らした確認すべき問い

Acceptance	データ収集・利活用について、 市民からどの程度まで受容されているか？
Benefit	データ利活用によって、どのような便益を、 誰が、何のために得られるのか？
Cost	コストは誰が、どの程度負担するのか？ Benefitに照らして適正か？
Data	どのようなデータ・データベース構造であるべきか？
Ecosystem	データの作り手や使い手、運用管理者など、 データを取り巻く関係者個々にとって 継続的な仕組みか？

(資料) みずほリサーチ & テクノロジーズが作成

くためである。一足飛びにデータベース構築には至らなくても、将来の接合可能性を見越して、表記ルール統一等の準備をしておくだけでも、構築上の大きな一歩になる⁽¹²⁾。

本稿では国内外の事例をもとに、データベース構築の検討を進める際に必要な点を検討した。構築したいデータ単体を議論対象とするのではなく、エコシステムの一構成要素としてのデータ(Data)として位置付け、データを取り巻くステークホルダーやその背景に目を向けることが重要である。本稿で紹介した「5つの視点」に照らした問いを紹介し、本稿を締めくくりたい。今後のEBPM推進に向けて、行政記録情報を活用したデータベース構築・整備の参考になれば幸いである。

注

- (1) 同事業の背景や日本の公的統計を取り巻く課題の一端は、萩野覚(2022)「統計の国際比較可能性の向上に向けて」『統計』第73巻10号にも詳しい。
- (2) 海外事例は総務省(2022)「公的統計の国際比較可能性に関する調査研究(社会統計編)報告書」を参照

されたい。

- (3) 例えば、総務省(2012)「諸外国における国民ID制度の現状等に関する調査研究報告書」によると、スウェーデンの教会による住民登録は1571年に開始され、1686年には登録方法が全国で統一されたことが報告されている。
- (4) 2022年10月時点で当社が確認したところ、スウェーデンのミュージシャン『ABBA』のメンバーや著名なジャズシンガーなども住所検索サイトによって住所や年齢などの閲覧が可能であった。
- (5) 保健福祉データ(Health and Welfare Data: HWD)センターにはNHIRDを含む70以上の医療関連データベースが管理されている(Hsieh et.al. 2018)。
- (6) 例えば、2019年に台湾では社会保険番号と各種の情報を統合した「デジタルIDシステム」を導入する動きがあった。このIDシステムは医療情報を一元化するシステムであり、導入に関して多くの議論が行われたが、市民党からの反発によって導入が見送られた。主な懸念点はデータがハッキングされ盗まれる可能性であったという。もちろん改善されれば今後、導入される可能性はあるが、2023年現在導入には至っていない。
- (7) 具体的には、「生活困窮判定」「学力判定」「非認知能力等判定」の3つの要素から、子どもに対する支援の可否が年2回判定されている。生徒情報が経年的に蓄積されているため、学力の絶対値だけではなく学力の変動から予兆が発見できる。また、主観的・定性的な側面での変化もいち早くキャッチアップできていることも特長である。実際、2017年後

半で判定の結果「重点支援」と判定された児童477人のうち、その半数弱の212名が学校では問題の予兆を認識していない児童であったという。

- (8) なお、現在ではNDBにとどまらず、更なるデータ拡充のためにデータベース間の連結が進められている。例えば、NDBと介護DB（介護レセプトと要介護認定情報を格納したDB）は2018年度以降のデータが連結され、2020年10月より研究目的に限り第三者提供が開始されている。また、匿名診療等関連情報（DPC データ）を格納したDPCDBについても、2022年のNDB及び介護DBとの連結に向けた取組みがなされている。
- (9) 前述の総務省事業で OECD 加盟国を中心とした計41の国や地域を対象とした調査を行ったところ、たしかに多くの国で個人番号が活用されているものの、ドイツやオーストラリアなど独自の方法で管理・構築している国も見られた。
- (10) ここでいう暗号化の代表例はハッシュ関数によるものである。ハッシュ関数とは要約関数とも呼ばれ、名前や生年月日などの任意の入力データから固定長の変数を乱数のように生成する関数である。そうした関数から得られた値はハッシュ値と呼ばれる。
- (11) 例えばNDBでも個人識別情報が匿名化処理されている。具体的には、レセプトに含まれる氏名や被保険者証の記号・番号、生年月日といった情報はデータベースに格納される際には削除される。その代わりにハッシュ関数によって匿名加工されたハッシュ値がデータに付与され、代替的な個人IDとされている。
- (12) なお、技術的にも課題はある。非常に低い確率ではあるものの、暗号化の特性として同一個人でないにもかかわらず同じIDが生成される可能性がある。これは仕様上避けることができない問題である。個人識別番号を用いないデータの構築にはこのように一定の不確かさを許容する必要があるといえる。

425-436.

5. 総務省(2012)『諸外国における国民ID制度の現状等に関する調査研究報告書』

参考文献

1. 総務省(2022)『公的統計の国際比較可能性に関する調査研究(社会統計編)報告書』
2. 森安亮介(2022)『見えない格差を可視化する、データの整備と活用例—教育分野を中心に—』みずほリサーチ & テクノロジーズ コンサルティングレポート vol.2
3. 栗山緋都美(2022)『一デンマーク型の行政サービスデザインからの示唆—行政のデジタル化』みずほリサーチ & テクノロジーズコラム
4. Hsieh, C. Y., Su, C. C., Shao, S. C., Lin, S. J., Kao Yang, Y. H., & Lai, E. C. C. (2018)「台湾の保健福祉データベース—ビッグデータ分析による高品質なリアルワールドエビデンスを得る機会—」『医薬品医療機器レギュラトリーサイエンス』Pharmaceutical and medical device regulatory science, 49 (7),