



RIETI Policy Discussion Paper Series 26-P-013

EBPM（エビデンスに基づく政策形成）入門 第6話 EBPMに必要なデータは役所の中に既にある

関沢 洋一
経済産業研究所



Research Institute of Economy, Trade & Industry, IAA

独立行政法人経済産業研究所

<https://www.rieti.go.jp/jp/>

EBPM(エビデンスに基づく政策形成)入門¹

第6話 EBPMに必要なデータは役所の中に既にある

関沢洋一（独立行政法人経済産業研究所）

要 旨

- ・ 行政機関が保有する業務情報は、網羅的であることや、長期にわたって同一人物のデータを追跡できることなど、政府統計やアンケート調査では得られにくいデータとしての価値があり、EBPMの推進や研究目的でその利用を推進することが重要である。
- ・ 諸外国では北欧諸国のように業務情報の利用が既に進んでいるところもあり、アメリカやイギリスも業務情報の利用を拡大する方向を模索している。
- ・ 業務情報の利用拡大は、プライバシー保護や行政機関の負担の増大への対応などの課題があり、相反する要請に向き合うことが重要になる。そのあり方についての基本的な考え方として「5つの安全」がある。
- ・ 地方公共団体は中央官庁よりも業務情報の利用拡大を進めていくポテンシャルが高く、業務情報を活用したランダム化比較試験の実施などの先駆的な取組が行われることが期待される。

キーワード：業務情報（行政記録情報）、縦断データ（パネルデータ）、デンマーク、5つの安全、地方公共団体

JEL classification: H11 H43

RIETI ポリシー・ディスカッション・ペーパーは、RIETI の研究に関連して作成され、政策をめぐる議論にタイムリーに貢献することを目的としています。論文に述べられている見解は執筆者個人の責任で発表するものであり、所属する組織及び（独）経済産業研究所としての見解を示すものではありません。

¹本稿の原案は、経済産業研究所（RIETI）のポリシー・ディスカッション・ペーパー検討会で発表を行ったものである。

第6話 EBPMに必要なデータは役所の中に既にある

(要約)

- ・ 行政機関が保有する業務情報は、網羅的であることや、長期にわたって同一人物のデータを追跡できることなど、政府統計やアンケート調査では得られにくいデータとしての価値があり、EBPMの推進や研究目的でその利用を推進することが重要である。
- ・ 諸外国では北欧諸国のように業務情報の利用が既に進んでいるところもあり、アメリカやイギリスも業務情報の利用を拡大する方向を模索している。
- ・ 業務情報の利用拡大は、プライバシー保護や行政機関の負担の増大への対応などの課題があり、相反する要請に向き合うことが重要になる。そのあり方についての基本的な考え方として「5つの安全」がある。
- ・ 地方公共団体は中央官庁よりも業務情報の利用拡大を進めていくポテンシャルが高く、業務情報を活用したランダム化比較試験の実施などの先駆的な取組が行われることが期待される。

政策介入の効果検証を行ってエビデンスを明らかにするためにはデータが必要です。データがなければどんなに優れた研究者でも効果検証は行えません。

あまり認識されていませんが、これらのデータの多くは役所の中に既にあります。出生届から死亡届に至る様々な届出、補助金の申請と採択、税務関係で提出する書類に記載された情報などです。本稿ではこれらを「業務情報 (administrative data)」と呼びます¹。

第1節では業務情報がEBPMにおいてなぜ重要なのか、第2節では諸外国でどのような取り組みが行われているのかについて述べます。第3節では「5つの安全」について述べ、第4節で日本の現状と課題について述べます。

第1節 業務情報の重要性とその利用を巡る課題

業務情報の特長

多くの国々が研究者による業務情報の利用に関心を有しており、研究者による利用を認めるようになってきました。業務情報にはアンケート調査や政府統計にはない次のような特長があります。

¹ 行政機関が保有する業務情報は行政記録情報と呼ばれることが多いのですが、ここでは健康保険組合が保有する健康診断やレセプトの情報など、行政機関以外が保有する情報で公的な側面の強いものも含めて業務情報と呼びます。

①観測数が多く、欠損値が少なく、偏りが小さい

業務情報以外のアンケート型の調査だと、国民全体を対象とする国勢調査や企業全体を対象とする経済センサスのような罰則担保で国が行っている調査でさえ回答率は7～8割で、把握できない部分が出てきます。

研究者自らが行うアンケート調査の場合はもっとうまくいかない場合が多く、調査票を送った人々のうち調査票に回答する人々が半数未満だったり、調査票自体には回答しても答えたくない部分を未記載にしたりする場合があります。特に所得や資産についての質問は答えない人々が多く、未回答によって数値の空白（欠損値）が生じやすくなります。所得税や法人税などの税務情報でこれらの統計調査を補えれば、偏りの少ない国民全体の様々な情報が得られることとなります。

このことを経団連が提言で指摘したことがあります。経済センサスについて、経済センサス活動調査の調査項目と、企業が法人税を申告する際に税務署に提出する法人事業概況説明書と一致している項目が多いことから、「法人事業概況説明書のデータを活用することで、経済センサス活動調査の調査事項代替や欠測値補完を行うことができ、報告者負担の軽減や統計精度の直接に寄与する」(P 8) と経団連の提言は指摘しています²。

②長期に渡って同一人物のデータを追跡しやすい

一時点における複数の個体（人々）のデータは横断データ（クロスセクションデータ）と呼ばれます。これに対して、同一個体のデータを複数時点に渡って観察するデータは縦断データ（パネルデータ）と呼ばれます。横断データよりも縦断データの方が時間による変化を分析に取り込めるので、因果関係を明らかにすることが求められる政策介入の効果検証においては断然有利です。

残念なことに、研究者が同一人物に対する複数回のアンケート調査などで研究参加者の情報を何年も追跡して縦断データを構築しようとしても、時間とともに未回答が増えていき、長期にわたる情報を得ることは難しくなります。また、縦断データを構築するに当たっては個人のID（識別情報）の厳密な管理が必要となり、研究者にとっての負担になります。こうした理由により、横断データと比べて縦断データの構築は難しいです。

ところが、業務情報を使えば質の高い縦断データを比較的容易に取得することが可能になります。例えば、個人の医療費をアンケート調査で把握しようとするならば、同じ人に調査票を毎年送っても安定的な回答はまず得られませんが、健康保険組合が保有している業務情報を使えば、個人毎の医療費の正確な推移が

² https://www.keidanren.or.jp/policy/2021/092_honbun.pdf

わかります。同じことは医療費に限らず、所得、定期健康診断を受けている人々の様々な数値（例えば血圧）など、いろいろな数値の推移についても言えます。

業務情報が研究目的で使えると何ができるようになるのか？

業務情報の研究者による利用を広げていくと、これまではわからなかったことがわかるようになったり、今まで行えなかった政策介入の効果検証が行えるようになったりします。以下で述べます。

①今までにわからなかった基礎的な情報がわかる

一例を挙げます。東京大学政策評価研究教育センター（CREPE）の「EBPM推進のための自治体税務データ活用プロジェクト」では、市町村の税務データを匿名化処理した上で CREPE に提供してもらい、研究に用いています³。この情報を用いて、いわゆる 103 万円の壁、130 万円の壁が実際にどの程度の就労抑制効果を有しているかが 2023 年の研究において明らかにされています⁴。

103 万円の壁のような重要なトピックであっても、これまでの研究はアンケート調査によるものが多かったために網羅的なデータに基づく信頼できる事実の把握が困難でした。例えば、103 万円の壁等を検証した 2003 年の研究では、所得を把握するためにアンケート調査形式の政府統計である国民生活基礎調査が使われていますが⁵、この調査は調査対象世帯数（約 30 万件）のうち約 3 割が回答していません（2022 年の場合）。

CREPE の行った取組は、役所の中に眠っていたデータを利用することによって今までにわからなかったことを明らかにした点で画期的です。

②様々な効果検証を行える

ここには 3 つの意味があります。第 1 に、個々の者がある政策介入を受けたか否か、どういう基準に基づいて政策介入が行われるかの情報は業務情報として政府内部に保管されています。この情報が研究者に提供されれば、これまでに行えなかった政策介入の効果検証ができるようになります。たとえば、中小企業への補助金の効果を検証するためには、補助金の採択・不採択それぞれの企業名や法人番号、採否の決定のために作成した企業ごとの得点についての情報が必要ですが、これらの情報は政府内部（あるいは政府から委託を受けた事業者など）に保管されています。

2 つめに、業務情報では縦断データが構築されて、医療費や所得などを長期に

³ 川口・正木（2022）

⁴ 近藤・深井（2023）

⁵ 大石（2003）

わたって追跡することができるため、ある時点で起きた出来事や政策介入を自然実験とみなして、その効果を検証することが可能になる場合があります⁶。たとえば、第5話で紹介した保健指導のRDD（回帰不連続デザイン）による効果検証は、健康保険組合が保有する健康診断の縦断データ（本稿ではこれは業務情報として扱います）を用いたものです⁷。産科病棟の閉鎖が出産関連行動に及ぼす影響についてのDID（差の差分分析）による効果検証（第3話で紹介したもの）では、アメリカの連邦政府が保有するNVSS（the National Vital Statistics System）のデータベースにアクセスして、1989～2019年の各郡の出生・死亡を始めとする業務情報を利用して分析を行っています⁸。

3つめに、業務情報を活用できるとRCT（ランダム化比較試験）の実施が容易になります。研究者としては個人のIDがわかる情報を研究参加者から取得して、それを政府が保有する業務情報と接合することができれば、自らアウトカムを計測しなくて済みます。後述します。

③データのリンケージにより幅広いことがわかる

多くの研究者（そしておそらくは多くの国民）がデータに関して知りたいのは、ある変数と別の変数の関係です。高学歴になると所得が増えるのか、所得が大きい方が健康なのか、といったものです。

複数の別々のデータセットが共通して持っている情報（個人を特定するIDなどで、マッチングキーと呼ばれます）を利用して、複数のデータセットを接続して1つのデータセットにすると、変数と変数の関係を明らかにできるようになり、データから読み取れる情報量が格段と増えます。このような接続を本稿では「リンケージ」と呼ぶことにします。

例えば、カナダでは教育水準や所得などの社会経済的地位と入院の関係について業務情報のリンケージにより検証が可能となっており、研究結果が報告されています⁹。それによると、教育水準の低さと所得の低さはそれぞれ独立して入院率の高さと関係しており、特に、ADSC（適切な外来診療を受けることで入院を回避できる疾患）においてはその傾向が大きいことが示されています。

業務情報の提供を巡る課題

ここまでの話を踏まえると、業務情報の利用を積極的に進めれば良さそうですが、そう簡単ではなく、課題がいくつもあります。

⁶ 別所他 (2019)

⁷ Sekizawa (2023); Sekizawa (2024)

⁸ Fischer et al. (2024)

⁹ Godley and Tang (2022)

1 つめはプライバシーの保護で、これが最も重要です。税務情報が典型ですが、個人や法人の所得をはじめとして、外部に流出させてはいけない情報があります。特に、様々なデータのリンケージが進んでいくと、たとえば、ある人がかかった病気を手がかりに、その人を特定し、所得を把握することができる場合があります。言い換えると、データのリンケージは個人のプライバシーを漏えいするリスクを高める面があります。

2 つめの課題は行政側の負担です。行政機関は業務情報を外部に提供することを想定して管理していないために、整理した形でデータを保管していない場合があります。提供用のデータセットを作るだけでも大変な作業です。多くの行政機関では情報提供のような突発的な業務増加に備えた人員配置も行っていないので、情報提供関連の業務が新たに生じると、ただでさえ忙しい行政官に残業を強いることになりかねません。また、データの処理は現場の能力を超えることも多く、時として外部の専門家の手伝いが必要になりますが、そこまでの予算はないことが多いです。以上に加えて、情報提供を受ける研究者がその情報を漏えいしないためにどのようにするかを検討する必要が生じ、こちらは漏えい防止のためのテクニカルな検討に加えて、契約や覚書の締結を含めて法的な考察も必要になります。

3 つめの課題として、研究目的で業務情報が研究者によって利用される場合、行政側として知られたくないことが明らかになる可能性があります。たとえば、提供する業務情報によって政策介入に効果がないという結果が明らかにされるのは行政側にとっては都合の悪いことが多いです（第7話参照）。また、研究者がデータをチェックする過程で間違いや不備が見つかって行政機関が外部から非難される可能性もゼロではありません。

第2節 海外における取組

業務情報を研究者に利用してもらって政策形成に役立つ貴重なエビデンスを明らかにするという要請と、個人のプライバシーを守ることや行政側の負担増を極力抑えることとの間のバランスをどうとるかについて、多くの国々が苦心してきました。以下では世界最高水準である可能性が高いデンマークの例を最初に取り上げ、その後で、アメリカ、イギリスの例を取り上げます。

世界の先端を行くデンマーク

他の北欧諸国と同様にデンマーク政府は高度な業務情報のデータベースを保有しています¹⁰。

1960年代の立法措置により各行政機関が保有する業務情報がデンマーク統計

¹⁰Mainz et al. (2019); Schmidt et al. (2015); 伊藤 (2020)

局 (Statistics Denmark) に提供されることが決まりました。また、国民一人一人に付与される ID 番号が作られて、各行政機関の情報のリンケージができるようになりました。

この結果として、個人 ID 番号をマッチングキーとして、生年月日、年齢、性別、所得、医療費、医療機関で診断された病名、死亡及びその原因といった様々な情報のリンケージを行うことが可能になっています。登録された研究機関に所属する登録された研究者が研究計画を明らかにして承認をもらうと、研究に必要な範囲で統計局を中心とする政府職員が ID 番号をマッチングキーとして複数のデータセットのリンケージを行った上で ID 番号を取り除き、研究に必要なデータセットができあがります。このデータに研究者は遠隔操作でアクセスできます。このデータを使って質の高い研究がいくつも行われています。

また、研究者が行う RCT のアウトカムの測定にこのデータベースが利用されています。研究者としては研究参加者の ID 番号さえ把握しておけば、デンマーク政府のデータベースに含まれているデータをアウトカム変数にして、介入終了後にそのアウトカム変数を利用することによって自らデータを取得しないで済みます。第5話で取り上げたデンマークの2つの大規模な健康診断の RCT では、政府が保有する死亡や重大疾患の業務情報が利用されています¹¹。

連邦政府という制約下でのアメリカの取り組み¹²

アメリカではデンマークとは異なって、各行政機関が保有する情報を1カ所に集約する方向には進んでいません¹³。連邦政府、州、郡といった多層的な行政機関があり、それぞれの独立性が強いので、データを中央集権的に管理することが難しいという事情があります。

地方政府のレベルでは業務情報を利用した政策効果の検証は多数行われており、ペンシルバニア大学の AISP (Actionable Intelligence for Social Policy) のように、こうした業務情報の利用拡大に向けた地方政府の取り組みをサポートする組織もできています¹⁴。

連邦政府のレベルでは NSDS(National Secure Data Service)という構想があり、プライバシーの保護を図りつつ、研究者によるデータのアクセスの改善、異なる行政機関が保有するデータのリンケージの増加が模索されています¹⁵。

2019年に成立した法律である Evidence Act (the Foundations for Evidence-

¹¹ Jørgensen et al. (2014); Lindholt et al. (2022)

¹² EBPM を巡るアメリカのデータの事情は津田・岡崎 (2018, 2022)で詳しく取り上げられています。

¹³ Hart and Potok (2020)

¹⁴ Zanti et al. (2022)

¹⁵ 津田・岡崎 (2018, 2022)

Based Policymaking Act of 2018)に基づいて、ACDEB (The Advisory Committee on Data for Evidence Building) が設立され、NSDS も含めてエビデンス構築のための連邦政府のデータの利用をどうやって進めるかについて、現状をレビュー・分析して、OMB (行政管理予算局) に対して提言することとなりました。ACDEB は 2 度の報告書を作成し、2022 年に出されたものが最終報告書になっています¹⁶。NSDS については 2022 年から 5 年間の予定でパイロットプログラムが進行中です¹⁷。

実際には、Evidence Act 法の制定以前から業務情報の利用を進めようとする動きが連邦政府内に見られており、こちらもまた重要だと思われるので、以下ではセンサス局と IRS (内国歳入庁) について取り上げます。

①センサス局¹⁸

アメリカのセンサス局は 1982 年に経済研究センター (Center for Economic Studies, CES) を設立し、縦断データのデータベースの開発と、秘匿性の高い情報を研究者が分析できる環境の整備を開始しました。研究者は詳細な研究提案書を提出し、その研究の実現可能性や必要性、情報の不正漏えいがないことを証明する必要があります。また、バックグラウンドチェックを受け、使用すべきデータセットや変数を明確にしたうえで、承認されたもののみアクセスが許可されます。

データの利用場所についても厳しく制限されており、全米に複数設置されている連邦統計研究データセンター (Federal Statistical Research Data Centers, FSRDC) でのみ利用が認められ、利用状況は常にモニターされています。さらに、データおよび研究成果物はセンサス局の職員による詳細なレビューを経て初めて公開が許可されます。

FSRDC のネットワークは、センサス局のデータにとどまらず、連邦政府の他の行政機関の秘匿性の高いデータも取り扱うようになってきています。また、2019 年からは一部のデータに対してリモートアクセスが開始されました。

最近の分析によると、センサス局の機密情報の利用は増加傾向にあり、このデータを使った学術研究が科学や政策における論争にもたらすインパクトは高まっているそうです¹⁹。ただし、このデータの利用は名声のある組織に所属する評価の高い研究者に概ね限られているとも指摘されています。

¹⁶ Advisory Committee on Data for Evidence Building (2022)

¹⁷ https://www.americasdatahub.org/wp-content/uploads/2024/11/NSDS_November_2024_Webinar_Slides.pdf

¹⁸ Nagaraj et al. (2025)

¹⁹ Nagaraj et al. (2025)

②内国歳入庁 (IRS)

IRS は徴税事務を通じて、アメリカ国民全体を網羅する所得等の巨大なデータを保有しています。法律によりこのデータを外部に提供することはできませんが、その制約の中で、研究目的や政策立案の参考情報として IRS が保有する情報を使った研究が進められてきました²⁰。IRS 内の統計部局は SOI (Statistics of Income) で、SOI は 2012 年に JSRP (Joint Statistical Research Program) を設立し、研究者との共同研究を行っています。

JSRP では研究者から提出された研究計画が吟味されて、限られた研究者が行う研究のみが機密データへのアクセスを認められます。その際に、研究者は厳しいバックグラウンドチェックを受けるとともに、主たる研究者は連邦政府の被雇用者として扱われて IRS 職員と同様の守秘義務と罰則の対象となります。SOI 職員が研究プロジェクトの監視役を務めます。

IRS のデータ利用は狭き門で、2023 年には 114 件の申請に対して選ばれたプロジェクトは 25 件でした。

IRS のデータを使った最近の研究例として、刑務所に収監されることがその後の雇用や収入にどのような影響を及ぼしたかについての効果検証を行ったものがあります²¹。この研究では、オハイオ州とノースカロライナ州の犯罪者のリストと IRS の所得等のデータのリンケージが行われています。マッチングキーとしては、社会保障番号 (SSN) や、氏名・誕生日・住所などの情報が使われています。分析の結果、刑務所に 1 年間収監されることによって 5 年間の累積所得が平均で 13%減ること、その後の雇用や賃金に影響するというエビデンスは得られなかったことが報告されています。

EBPM 発祥国であるイギリスにおける取り組み²²

イギリスでは、Statistics and Registration Service Act (SRSA) 2007 により、国家統計局 (Office for National Statistics) が保有する未公表データについて、統計研究目的で承認を受けた研究者がアクセスできるようになりました。さらに、Digital Economy Act (DEA) 2017 によって、政府内外の研究者コミュニティを支援する法的枠組みが構築され、承認を受けた研究者が行う公益目的を有する研究プロジェクトのために、個人が特定できないように匿名化された行政記録情報を提供することが可能になっています。

DEA および SRSA に基づき、研究認定委員会 (RAP) が研究者、研究プロジェクト、認定処理機関の認定及び監督を行っています。承認された研究者のプロ

²⁰ Burman et al. (2024); Mervis (2014); O'Hara et al. (2024)

²¹ Garin et al. (2025)

²² UK Data Service (2025); UK Statistics Authority (2025a, 2025b)

プロジェクトが認定されると、RAPによる審査を経て認定された処理機関（2025年4月時点で12機関）が提供する安全な環境の下で、匿名化された業務情報の利用が可能になります。

また、イギリスでは研究者と行政機関の間に入って業務情報の研究目的での利用を促進するためのパートナーシップとしてADR UK（Administrative Data Research UK）が設立され、予算も付与されています²³。

次節で言及する「5つの安全」という枠組みは2003年にイギリスの国家統計局で登場したもので²⁴、イギリスが早くから業務情報の利用拡大と個人情報の保護のバランスをどう図るかを模索してきたことを示唆します。

第3節 5つの安全

研究目的で業務情報の利用拡大を進めることと個人情報を保護することはトレードオフの関係にあり、両者の緊張関係の解決はどの国でも課題です。

この問題の解決に向けて、5つの安全という枠組みが提示されています²⁵。5つの安全とは、安全な人々（safe people）、安全なセッティング（safe settings）、安全なデータ（safe data）、安全なアウトプット（safe output）、そして、安全なプロジェクト（safe projects）を指します。

安全な人々

データの漏えいを防ぐためには、漏えいしそうにない人々だけにデータへのアクセスを許可することが重要になります。

第一に、データを利用する人々に一定の能力と資質が求められます。統計学やデータサイエンスなどの習得を通じてデータ処理能力があること、機密情報を扱うための倫理関係の講習を受けていること、犯罪歴がないことなどが考えられます。

第二に、安全な組織（safe organization）ということが挙げられます。大学を始めとする研究機関に所属する研究者に限定して、研究者のみならずその所属研究機関側もデータ提供過程に関与させることによって、大学にも責任を分担させることが考えられます。

第三に、データにアクセスする研究者を非常勤の公務員とすることによって「安全な人々」になってもらう場合もあります。上述したIRSの例がこれに該当します。主たる研究者は連邦政府の被雇用者として扱われてIRS職員と同様の守秘義務と罰則の対象となります。日本でも似た例があり、後述します。

²³ Waind (2020)

²⁴ Desai et al. (2016)

²⁵ Desai et al. (2016); 伊藤 (2016); 岡田 (2023)

安全なセッティング

これはデータへのアクセスが許容される環境についてのものです。一番厳密な対応としては、データ提供側が設定した特定の場所と特定の情報機器を使ってデータにアクセスすることとして、更に、データへのアクセスが監視されているようにすることが考えられます。たとえば、後述する国税庁における税務情報のアクセスにおいては、税務大学校内における同庁の情報機器においてのみデータがアクセスできるようになっています。

もう少し緩くなると、上述したアメリカのセンサス局の FSDRC のように、国内の数カ所（主として大学）にデータにアクセスできる場所を設ける場合があります。政府統計については日本でも実例があります。

もう少し緩めていくと、遠隔アクセスで外部から業務情報を利用することが可能な場合があります。デンマークの情報アクセスはこの例です。

安全なデータ

データそのものを安全にすることによって仮にデータが漏えいしてもリスクを最小限にとどめることが考えられます。

最低限の話として個人の ID をデータから取り除くことがあります。ただし、この前提としてデータのリンケージを誰かが既に行っていることが必要な場合が多いです。たとえば、デンマークの業務情報に研究者がアクセスする場合には、健康・教育・所得など様々な情報を一緒に使うことによって、たとえば、教育の健康や所得への影響を調べることができるわけですが、通常はこれらの業務情報は別々の行政機関が保有しています²⁶。この業務情報をデンマーク統計局が行政機関から集めて個人の ID をマッチングキーとしてデータのリンケージを行い、その後で個人の ID をデータセットから取り除くことによって、個人の ID が取り除かれたデータが完成します。

残念なことに、個人の ID が取り除かれたデータセットであっても、いくつかの変数を用いることで個人を特定できる可能性があります。たとえば、郵便番号・生年月日・性別・子供の数など 15 の変数があれば、ほとんどのアメリカ人の健康情報が特定されるのだそうです²⁷。

この課題への対応策としては様々な方法が提案されています²⁸。たとえば、統合データ (synthetic data) というデータがあり、これは元のデータセットをそ

²⁶ 伊藤 (2020)

²⁷ Rocher et al. (2019)

²⁸ 専門的で詳しい話は伊藤 (2022)に書かれています。

の統計的な意味を損なうことない別のデータセットに置き換えます²⁹。ただし、オリジナルのデータと統合データで分析結果が異なる場合もあるので、統合データは完全な解決策とは言えません。

それでも、研究者が統合データを使って予行演習を行って統計ソフトの分析コードを作成し、その分析コードを行政機関内の業務情報の管理者に渡してそのコードで統計ソフトを動かしてもらい、できあがった分析結果を研究者が受け取れるようになれば、研究者が業務情報本体にアクセスしなくても分析ができるようになります。

安全なアウトプット

分析結果を公表する際に個人情報が出ないための取り組みを指します。通常はデータ提供側がチェックします。個人の名前が出ないことはもちろんですが、データ提供主体の名前（市町村名など）を出さないことが求められることもあります。

安全なプロジェクト

提案されたプロジェクトが機密データにアクセスして分析する上で適切かどうかをみるもので、5つの安全の中で最も重要とされます³⁰。当該データを適切に処理できる分析者による研究体制が整えられているか、行政側の負担を強いてでも行うに値する公益性の高い研究と言えそうか、データを安全に管理できる状態になっているかなどが研究計画書によって明らかにされ、データ提供側の審査に耐えられるものである必要があります。諸外国の例では、大学などの倫理審査委員会が研究計画を審査してプロジェクトを承認することをデータ提供の条件にする場合があります³¹。日本でも厚生労働省が管理する医療データベース（NDB）の利用に当たって倫理審査委員会における承認が必要とされています。

5つの安全は業務情報を安全に利用するという目的を実現するための一つのアプローチであり、絶対的なものではありません³²。また、技術の進歩などにより時間とともに対応策が変化していくことにも留意する必要があります。

第4節 日本のにおける業務情報の提供を巡るいくつかの論点

²⁹ Kokosi et al. (2022)

³⁰ Green and Ritchie (2023)

³¹ Cole et al. (2020)

³² Green and Ritchie (2023)

日本でも研究目的での業務情報へのアクセス拡大の重要性を主張する人々は研究者を中心に増えています³³。日本学術会議も「行政記録情報の活用に向けて」という提言を2020年9月に発表しています³⁴。

ここから先は国と地方政府に分けて議論を進めます。

業務情報の利用拡大に向けた取り組み（国の場合）

業務情報の利用拡大に向けた主な取り組みとして、厚生労働省による医療関係データ（NDB）と、財務省による取り組みがあります。

NDBは健康保険組合が保有するレセプトデータに含まれる情報（診断名、処方した薬、医療費など）やメタボ健診の情報を厚生労働省が収集して匿名化し、研究目的での利用に供するものです³⁵。

財務省は、輸出入申告データと税務データを研究者に利用して研究してもらう取り組みを2021年から開始しました³⁶。

このうち税務データについては、国税庁保有行政記録情報を利用して税務大学校職員と共同で研究者が研究を行う制度が運用されています。学術研究目的やEBPM目的で研究者が行う研究について、公募に応募する形で研究計画を提案して、有識者会議の承認を得られると、その研究者は税務大学校の客員教授になって、同校の職員との共同研究として、国家公務員法第100条等の守秘義務を守り、個票データの利用を税務大学校の施設内で行い、研究ができます³⁷。

今のところ以上のような税務情報の利用は様々なファクトを明らかにすることが中心になっていますが、今後はEBPMの中心となる政策介入の効果検証にも使われることが期待されます。

たとえば近年著しく増えた企業への補助金に効果があるかというテーマがあります。第3話で紹介したチェコ共和国の研究開発補助金のように企業向けの補助金の効果検証が世界的にいくつか行われています。日本でも企業への補助金が増えています。効果検証となる研究は十分に行われていません。

その最大の理由として業務情報を分析に利用できないことがあります。補助金の採択に関する業務情報は経済産業省内にあり、アウトカム変数となる企業の付加価値額のデータは法人税申告時の企業からの申請情報として国税庁内にあります。補助金の効果を検証するためには、経産省内にある補助金情報と、国

³³ 近藤 (2024); 川口・正木 (2022); 中室 (2015); 田中 (2022); 藤澤他 (2022); 別所他 (2019)

³⁴ <https://www.scj.go.jp/ja/info/kohyo/kohyo-24-t297-3-abstract.html>

³⁵ 鈴木 (2025)

³⁶ 小川 (2025); 北村 (2025)

³⁷ 国税庁 (2024); 佐藤 (2025)

税庁にある法人の付加価値額のデータの両方が必要で、法人番号をマッチングキーとして2つのデータセットのリンケージを行うことが必要になります。そこまではまだたどりつけていません。

個人のプライバシーや法人の機密情報を守ることは重要なことではあり、やむを得ない面はあるのですが、他の国の先進事例と比べると更に先に進む余地があると思います。

更に先に進むためにどうすべきか

研究目的での業務情報の利用を進める上での主な課題として、プライバシーの保護の他に、①行政の現場の負担が増える、②行政機関にとって不都合な真実が明らかになる場合がある、③提供していいかどうかの判断について現場の行政官ではリスクと責任を負えないことがあります。

これらに向き合うための1つのあり方として、政府内に業務情報の利用を巡る諸課題に対応する部局（以下では「業務情報対応部局」）を作ることが考えられます。業務情報対応部局が中心となって、業務情報提供のためのガイドラインを作ったり、現場の行政官にリスクと責任を負わせないようにするための取り組みが進められたりすると望ましいです。

業務情報対応部局には、行政機関の担当部局と研究者の間に入って、どこまでの範囲の情報を提供するか、セキュリティ確保をどうやって図るかなどについて両者の間を調整することが望まれます。業務情報の提供要求に行政機関の個々の部局は不慣れな場合が多いですが、業務情報対応部局に対応ノウハウが蓄積されていれば、技術的事項や法的事項について現場の行政官の負担を軽減することが期待されます。

また、業務情報対応部局に期待される役割として、研究者からの申請を受けて必要な対応を個々の行政機関の担当部門に求めることがあります。ただ、無限定の申請があれば業務情報対応部局も個々の行政機関の担当部門もパンクしてしまうので、何らかのブレーキも必要になります。上述の国税庁の事例にあったような有識者会議のような場を使って検討するというのもあるかもしれません。

地方公共団体の果たす役割

日本の中央官庁は各行政機関の独立性が高く、デンマークのように各行政機関が保有する情報がリンケージされることには高いハードルがあります。日本同様に中央政府の行政機関の独立性の高いアメリカにおいては、連邦政府よりも州やその下の地方政府において、業務情報の利用拡大が進んでいます³⁸。

地方公共団体は中央官庁と異なって政策領域毎に別々の組織となっていない

³⁸ Zanti et al. (2022)

ことが一つの特徴であり³⁹、また、首長がリーダーシップを発揮しやすいので⁴⁰、中央官庁よりも地方公共団体の方が業務情報の利用拡大を進めていくポテンシャルが高いです。個々の市町村で見れば人口が少なくても分析がしにくくても複数の市町村の連携ができればビッグデータに近づいていきますし、政令指定都市のような大都市であれば最初からビッグデータになっています。

業務情報の研究目的での利用は医療・健康と教育を中心にいくつかの地方公共団体で既に取り組みが行われています。ただ、因果関係を特定できる RCT のような実験的な取り組みはほとんど行われておらず、観察データに基づくものが中心になっています。

今後期待されるのは、市町村の業務情報を利用して RCT が行われることです。

医療・健康関係を例にすると、市町村には国民健康保険の情報と後期高齢者医療関係の情報が集まり、また、死亡情報の届け出は市町村に対してなされます。更に、介護保険制度の運営者として、市町村は介護保険の被保険者情報や、介護サービスの提供状況に関する情報も保有しています。市町村から外部に流出させることなく、これらの情報を役所内で必要な範囲でのみリンケージすることができれば、数百人以上の規模で RCT を行える可能性があります。

たとえば、地方公共団体が行う後期高齢者を対象とした健康プログラム（認知症やフレイルを予防するための運動や食事や生活習慣の管理などの取り組み）の数百人規模の効果検証を RCT や RDD によって実施できるかもしれません。この場合、市町村が保有する情報とプログラム参加者の ID とのリンケージが可能になれば、要介護認定、医療費、健康寿命を損なう重要疾患（脳卒中）の発症など様々なアウトカム変数を設定することができます。

数少ない例として、大阪府寝屋川市で行われた RCT では、介護保険で「要支援」の認定を受けた高齢者 375 名をランダムに 2 群に分けて、両群に対して標準的なケアを提供するとともに、介入群に対してのみ、追加的にリハビリテーション専門職が主導する短期集中型自立支援プログラムを提供しました⁴¹。介入群では 190 名中 21 名（11.1%）が「要支援」から脱した一方で、対照群では 185 名中 7 名（3.8%）で、「要支援」から脱する割合は介入群の方が有意に高くなっていました。

地方公共団体で行われる RCT は、他の地方公共団体や国全体にとって重要な情報を提供する一方で、実施主体自体にとっては実施のメリットが大きくなく、本来望まれるほど実施されないことが想像されます。この点に対処して、RCT を始めとする先駆的な取組が地方公共団体で行われる事例を増やすためには、

³⁹ 曾我 (2019)

⁴⁰ 別所他 (2019)

⁴¹ Hattori et al. (2019)

補助金や国からの委託のようなインセンティブの付与を通じて、チャレンジする地方公共団体を国が後押しすることが必要なのかもしれませんが。

終わりに

EBPMを進めていく上で必要な業務情報の利用拡大とそれに伴う個人のプライバシーの漏えいリスクの高まりはトレードオフの関係にあり、容易に解決できるものではありません⁴²。

総論としてEBPMの推進を支持する人々が多いと思いますが、各論に入っていくとどうしても業務情報を巡る課題に向き合わざるを得なくなります。デンマークのようにデータ利用に重点を置く国、アメリカやイギリスのようにどのような方向に進むべきか模索している国など様々ですが、日本の場合にはそういう課題があるという認識があまり強くないように思われ、まずは課題を認識した上で悩むところからスタートする必要があるように思います。

引用文献

- Advisory Committee on Data for Evidence Building. (2022). *Advisory Committee on Data for Evidence Building: Year 2 Report*. <https://www.bea.gov/sites/default/files/2022-10/acdeb-year-2-report.pdf>
- Bowen, C. M. (2024). "Government Data of the People, by the People, for the People: Navigating Citizen Privacy Concerns," *Journal of Economic Perspectives*, 38(2), 181–200.
- Burman, L., Johnson, B., Bryant, V. L., MacDonald, G., & McClelland, R. (2024). "Protecting Privacy and Expanding Access in a Modern Administrative Tax Data System," *National Tax Journal*, 77(4), 927-947.
- Cole, S., Dhaliwal, I., Sautmann, A., & Vilhuber, L. (2020). Using Administrative Data for Research and Evidence-Based Policy: An Introduction. In S. Cole, I. Dhaliwal, A. Sautmann, & L. Vilhuber (Eds.), *Handbook on Using Administrative Data for Research and Evidence-based Policy*.
- Desai, T., Ritchie, F., & Welpton, R. (2016). "Five safes: designing data access for research," *Economics Working Paper Series*, 1601, 28.

⁴² Bowen (2024)

- Fischer, S., Royer, H., & White, C. (2024). "Health Care Centralization: The Health Impacts of Obstetric Unit Closures in the United States," *American Economic Journal: Applied Economics*, 16(3), 113–141.
- Garin, A., Koustas, D., McPherson, C., Norris, S., Pecenco, M., Rose, E. K., . . . Weaver, J. (2025). "The Impact of Incarceration on Employment, Earnings, and Tax Filing," *Econometrica*, 93(2), 503-538.
- Godley, J., & Tang, K. L. (2022). "Income, education, and hospitalization in Canada: results from linked census and administrative data," *Discover Social Science and Health*, 2(1), 19.
- Green, E., & Ritchie, F. (2023). "The present and future of the Five Safes framework," *Journal of Privacy and Confidentiality*, 13(2).
- Hart, N., & Potok, N. (2020). *Modernizing U.S. Data Infrastructure: Design Considerations for Implementing a National Secure Data Service to Improve Statistics and Evidence Building*. <https://ssrn.com/abstract=3700156>
- Hattori, S., Yoshida, T., Okumura, Y., & Kondo, K. (2019). "Effects of Reablement on the Independence of Community-Dwelling Older Adults with Mild Disability: A Randomized Controlled Trial," *International Journal of Environmental Research and Public Health*, 16(20), 3954.
- Jørgensen, T., Jacobsen, R. K., Toft, U., Aadahl, M., Glümer, C., & Pisinger, C. (2014). "Effect of screening and lifestyle counselling on incidence of ischaemic heart disease in general population: Inter99 randomised trial," *BMJ*, 348, g3617.
- Kokosi, T., De Stavola, B., Mitra, R., Frayling, L., Doherty, A., Dove, I., . . . Harron, K. (2022). "An overview of synthetic administrative data for research," *Int J Popul Data Sci*, 7(1), 1727.
- Lindholt, J. S., Søgaard, R., Rasmussen, L. M., Mejldal, A., Lambrechtsen, J., Steffensen, F. H., . . . Diederichsen, A. C. P. (2022). "Five-Year Outcomes of the Danish Cardiovascular Screening (DANCAVAS) Trial," *New England Journal of Medicine*, 387(15), 1385-1394.
- Mainz, J., Hess, M. H., & Johnsen, S. P. (2019). "The Danish unique personal identifier and the Danish Civil Registration System as a tool for research and quality improvement," *International Journal for Quality in Health Care*, 31(9), 717-720.
- Mervis, J. (2014). How Two Economists Got Direct Access to IRS Tax Records.

Science Magazine.

- Nagaraj, A., Stipanovic, F., & Tranchero, M. (2025). "Restricted-Access Data in Economics: Adoption, Diffusion, and Impact of U.S. Census Bureau's Microdata."
- O'Hara, A., Straus, S., Borzekowski, R., Arnsberger, P., & Johnson, B. (2024). "A Secure Query System to Improve Access to Individual Income Tax Data," *National Bureau of Economic Research Working Paper Series*, No. 32969.
- Rocher, L., Hendrickx, J. M., & de Montjoye, Y.-A. (2019). "Estimating the success of re-identifications in incomplete datasets using generative models," *Nature Communications*, 10(1), 3069.
- Schmidt, M., Schmidt, S. A., Sandegaard, J. L., Ehrenstein, V., Pedersen, L., & Sørensen, H. T. (2015). "The Danish National Patient Registry: a review of content, data quality, and research potential," *Clin Epidemiol*, 7, 449-490.
- Sekizawa, Y. (2023). "Effects of being eligible for specific health guidance on health outcomes: A regression discontinuity analysis using Japan's data on specific health checkups," *Prev Med*, 172, 107520.
- Sekizawa, Y. (2024). "Japan's intensive health guidance program has limited effects on cardiovascular risk factors: a regression discontinuity analysis," *Public Health*, 232, 108-113.
- UK Data Service. (2025). UK Data Service accredited for the provision and preparation of data under the Digital Economy Act. <https://ukdataservice.ac.uk/2025/04/21/digital-economy-act/>
- UK Statistics Authority. (2025a). "Accessing Data via the Research Accreditation Framework."
- UK Statistics Authority. (2025b). List of Digital Economy Act Accredited Processing Environments. <https://uksa.statisticsauthority.gov.uk/digitaleconomyact-research-statistics/better-access-to-data-for-research-information-for-processors/list-of-digital-economy-act-accredited-processing-environments/>
- Waind, E. (2020). "Trust, security and public interest: striking the balance A narrative review of previous literature on public attitudes towards the sharing, linking and use of administrative data for research," *Int J Popul Data Sci*, 5(3), 1368.

- Zanti, S., Berkowitz, E., Katz, M., Nelson, A. H., Burnett, T. C., Culhane, D., & Zhou, Y. (2022). "Leveraging integrated data for program evaluation: Recommendations from the field," *Evaluation and Program Planning*, 95, 102093.
- 伊藤伸介. (2016). "諸外国における政府統計マイクロデータの提供の現状とわが国の課題," *中央大学経済研究所年報*, 48, 233-249.
- 伊藤伸介. (2020). "デンマークとオランダにおける医療健康データの二次利用について," *日本統計学会誌*, 50(1), 109-138.
- 伊藤伸介. (2022). "大規模データにおける匿名加工と統計情報としての提供可能性," *評価クォーターリー*(62), 2-17.
- 岡田法大. (2023). "ゲノム情報を研究利用する際の情報の流れ," *医薬産業政策研究所政策研ニュース= OPIR views and actions*(69), 39-48.
- 近藤絢子. (2024). "税務データの研究利用可能性と EBPM," *とよなか都市創造*, 2, 23-29.
- 近藤絢子・深井太洋. (2023). "市町村税務データを用いた既婚女性の就労調整の分析 " *RIETI ディスカッションペーパー*, 23-J-049.
- 国税庁. (2024). *税務大学校との共同研究における国税庁保有行政記録情報利用に係るガイドライン*.
<https://www.nta.go.jp/about/organization/ntc/guideline.pdf>
- 佐藤黎. (2025). "税務情報の利活用—税務大学校との共同研究と匿名データの提供—," *統計*, 特集 税務情報のマイクロデータ研究(2025年9月号), 4-10.
- 小川英治. (2025). "序文," *フィナンシャル・レビュー*, 160, 1-4.
- 川口大司・正木祐輔. (2022). 「CREPE によるプロジェクト設立の背景とねらい」連載 「行政データと実証経済学: 東京大学 CREPE 自治体税務データ活用プロジェクトの実践」 第 1 回. *経済セミナー 2022 年 6・7 月号*.
- 曾我謙悟. (2019). *日本の地方政府: 1700 自治体の実態と課題*. 中央公論新社.
- 大石亜希子. (2003). "有配偶女性の労働供給と税制・社会保障制度," *社会保障研究*, 39(3), 286-300.
- 中室牧子. (2015). 「学力」の経済学. ディスカヴァー・トゥエンティワン.
- 津田広和・岡崎康平. (2018). "米国における Evidence-based Policymaking (EBPM) の動向," *RIETI ポリシー・ディスカッション・ペーパー*, 18-P-016.
- 津田広和・岡崎康平. (2022). *米国における EBPM*. 大竹文雄・内山融・小林庸平 編著 *EBPM: エビデンスに基づく政策形成の導入と実践*. 日本経済新聞出版.

- 田中隆一. (2022). "自治体行政データ," *日本労働研究雑誌*, 64(4), 57-60.
- 藤澤啓子・深井太洋・広井賀子・中室牧子. (2022). "就学前における家庭の経済的困難及び児童の健康・発達面のリスクと学力との関連:行政記録情報による検証," *発達心理学研究*, 33(4), 332-345.
- 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*, 141, 106-119.
- 北村行伸. (2025). "序文:日本における税務データを用いた近年の研究動向と今後の展望," *フィナンシャル・レビュー*, 160, 92-95.
- 鈴木里彩. (2025). "匿名医療保険等関連情報データベース (NDB) の二次利用に係る最近の動向," *統計*, 特集 行政情報のマイクロデータ研究(2025年8月号), 4-13.