

EBPM推進のための自治体税務データ活用プロジェクト

令和5年（2023年）5月

趣旨

デジタル改革関連法の施行を見据え、複数の地方自治体及び東京大学政策評価研究教育センター（CREPE）が連携し、①政策現場におけるEBPM（evidence-based policy making）の推進と②アカデミアにおける実証研究の発展を目指す。

背景

地方行政のデジタル化・スマート自治体化が主要な政策課題に

国や自治体におけるEBPMの推進

AI（人工知能）等のビッグデータの分析技術の発展

2000年代以降、行政データを用いた実証経済学が発展

目的

① 政策現場におけるEBPMの推進

- ・必ずしもリソースに余裕がない自治体も含めて、最新の経済学やAIに基づくデータ分析を利用可能にする。
- ・法学者・情報工学者の参画のもと、個人情報保護と両立したデータ利活用を実現

② アカデミアにおける実証研究の発展

- ・匿名化されたマイクロデータを基に研究を実施

概要

- ・自治体が、匿名化された個人レベルの税務情報をCREPEに提供
- ・CREPEが、計量経済学の知見やAI（人工知能）を用いて税収予測等のデータ分析を行い、参加自治体にフィードバック
- ・CREPEが、提供されたデータに基づき学術研究を実施

概要

募集対象自治体

- 全ての市区町村（※）
※ 2023年度は、都道府県の募集は行っておりません。
- 参加いただいた自治体には、計量経済学の知見を用いた2024年度の個人住民税の税収予測を提供する予定です。性別や年齢等の属性や、所得割・均等割等の区分に応じた記述統計表も提供いたします。

参加費

- 無料

応募要件

- 提供いただいたデータに基づく研究結果の公表（※）に同意すること。
※学術論文、研究報告資料、ディスカッション・ペーパー、一般向け記事等の研究成果物の公表を想定しております。なお、自治体の同意がない限り自治体名を明らかにすることはありません。研究結果の公表についての考え方は、自治体データプロジェクトにおけるデータ取扱規則第8条を参照ください。
- 次ページの必須データを匿名化（※）した上で、CREPEに提供できること。
※ 匿名化については、CREPEが匿名化のツールを配布するため、匿名加工技術の知識やプログラミングの知識は不要です。

応募要件

必須データ項目

少なくとも以下を含む（※1）個人単位のデータ（※2）

1. 年を示す変数(5年分以上)
2. 宛名番号（※3）
3. 収入金額：
営業等、農業、不動産、利子、配当、給与、給与（専従）、雑所得（公的年金）、雑所得（その他）

給与、給与（専従）、雑所得（公的年金）**以外**については、所得金額があれば、収入金額がなくても応募可能とします。
その他、そもそもシステムで保持していないデータ項目がある場合は、御相談ください。
4. 控除の金額：基礎控除、配偶者控除、配偶者特別控除、扶養控除、寡婦控除、ひとり親控除、勤労学生控除、社会保険料控除、生命保険料控除、地震保険料控除、小規模企業共済等掛金控除、雑損控除、寄付金控除、控除合計
5. 総所得金額、合計所得金額、総所得金額等
6. 住民税額：均等割、所得割、住民税額合計（利子割、配当割、株式等譲渡所得割を含んだもの）
7. 属性に関する項目：
生年月、性別、世帯番号、続柄、死亡年月日、転出年月日、住民になった日、住民区分（住登外納税者や死亡者等、1月1日時点の住民以外の者を識別するための変数）、勤務先を示す識別番号（※4）

※1 ここに示す以外のデータ項目も、データ分析において有用である可能性があるため、できる限り幅広く御提供いただけると幸いです。

※2 年を示す変数・宛名番号・属性に関する項目については、納税の有無にかかわらず当該自治体の全住民を含むものをご提供ください。その他の変数については、全課税者が含まれていれば全住民を含んでいる必要はありません。

※3 自治体独自に各個人に一意に割り振った番号のことであり、「個人番号」いわゆる「マイナンバー」とは異なります。提供に当たっては、世帯番号とともに、ハッシュ化（p. 10Q9及び（参考資料1）個人情報保護制度との関係及び匿名化処理）参照）していただきます。

※4 自治体独自で管理されている番号で問題ございません。

個人情報保護・匿名化

- 自治体様には、匿名化したデータをCREPEに提供いただくこととしています。本プロジェクトでは、自治体様の多様なニーズに応えるため、個人情報保護及び匿名化の専門家の監修の下、以下の匿名化手法を御用意しています（※1）。

匿名化手法

0. 氏名、個人番号（いわゆるマイナンバー）を事前に削除。
 1. 宛名番号、世帯番号、勤務先を示す識別番号のハッシュ化
 2. 宛名番号をキーにして、複数年のデータ結合
 3. 生年月日の月単位への丸め
- 匿名化方法について確認いただいた後、CREPEから自治体様に、R（※2）による匿名化プログラムを配布しますので、CREPEが配布するマニュアルに従って、自治体様のPCにおいて実行いただきます。
 - 併せて、CREPEにおいても、個人識別行為の禁止等を定めた「自治体税務データ活用プロジェクトにおけるデータ取扱規則」、「自治体税務データ活用プロジェクトにおける安全管理措置等措置等に関する規程」を定めており（※3）、個人情報の管理に万全を尽くしています。

※1 詳細は「（参考資料1）個人情報保護制度との関係及び匿名化処理」を参照ください。

最終的に、どこまでの匿名化が必要十分かについては、各自治体において判断いただくこととなりますが、その際、これらの資料を参考にしてください。

※2 学術的な統計分析に広く用いられている無料のプログラミング言語です。

※3 https://web.iss.u-tokyo.ac.jp/jichitai_data/data.html

想定スケジュール（案）

2023年5月	東京大学	プロジェクトへの参加を希望する自治体を募集します。
6月～7月	東京大学	協力自治体を選定します。
7月～8月	協力自治体	保有する税務データのデータ項目・定義を東京大学にお伝えいただきます。 この時点で自治体側にデータを抽出していただき、必須データに関して変数名等を確認していただきます。また、要配慮個人情報（※1）が含まれている場合は、変数名と定義を確認していただきます。
8月～12月	東京大学	協力自治体の提供データの概要（データの形式・分布）を確認します。
	東京大学	匿名加工の具体的手順書を作成し、協力自治体に送付します。
	協力自治体	手順書に沿って税務データを匿名加工し、東京大学に送付します。
	10月～ 2024年2月	東京大学

※1 要配慮個人情報とは、個人情報保護法2条3項に規定する「要配慮個人情報」と個人情報保護法60条5項に規定する「条例要配慮個人情報」のことを指します。これらの変数につきましては、研究において必要があるときのみ提供をお願いします。

※2 以上のスケジュールは現段階の想定であり、進捗に応じて、変更する場合があります。

研究体制

総括班

川口 大司
北尾 早霧
近藤 絢子
古川 知志雄
正木 祐輔

所得リスク・格差班（税務データを活用した所得リスクと所得格差の分析）

- 北尾 早霧 東京大学大学院経済学研究科 教授
独立行政法人経済産業研究所 上席研究員（特任）
- 鈴木 通雄 東北大学大学院経済学研究科 准教授
- 山田 知明 明治大学商学部 教授

雇用・社会保障班（セーフティネットと雇用・家庭）

- 近藤 絢子 東京大学社会科学研究所 教授
- 深井 太洋 筑波大学人文社会系 助教
東京大学政策評価研究教育センター 招聘研究員

税理論・実験班（最適税制論理論モデルと徴税率フィールド実験）

- 古川 知志雄 横浜国立大学大学院国際社会科学研究院・経済学部 准教授
東京大学政策評価研究教育センター 招聘研究員
- 別所 俊一郎 早稲田大学政治経済学部 教授

学術利用基盤整備班（個人情報保護、匿名化を含めた行政記録情報の学術利用基盤整備）

- 川口 大司 東京大学大学院経済学研究科 教授
東京大学公共政策大学院 副院長・教授
東京大学政策評価研究教育センター 前センター長
- 佐藤 一郎 国立情報学研究所 教授
- 穴戸 常寿 東京大学大学院法学政治学研究科 教授
- 正木 祐輔 東京大学政策評価研究教育センター 招聘研究員
独立行政法人経済産業研究所 コンサルティングフェロー
（総務省から神戸市に派遣中）

※ 2023年5月時点

※ 「○」は班長

※ 教育データが得られた場合など、必要に応じて、その他の研究者も参画予定

個人情報保護等に関する法的検討体制

- 本プロジェクトでは、匿名化を行っているものの、個人単位の自治体データを扱うため、個人情報保護等に万全を期す必要があります。
- 学術利用基盤整備班では、法学者を中心とする法学班を設け、個人情報保護法や地方税法をはじめとする法的検討を専門的に行っています。

学術利用基盤整備班（個人情報保護、匿名化を含めた行政記録情報の学術利用基盤整備）

○ 川口 大司 佐藤 一郎 穴戸 常寿 正木 祐輔

統括補佐

細田 幸恵

法学班

	小川 亮	東京都立大学法学部	助教
	神山 弘行	東京大学大学院法学政治学研究科	教授
○	穴戸 常寿	東京大学大学院法学政治学研究科	教授
	巽 智彦	東京大学大学院法学政治学研究科	准教授
	平田 彩子	東京大学大学院法学政治学研究科	准教授
	藤谷 武史	東京大学社会科学研究所	教授
	船渡 康平	信州大学経法学部	准教授
	正木 祐輔	東京大学政策評価研究教育センター	招聘研究員
	山羽 祥貴	東京都立大学法学部	准教授

※ 2023年5月時点

※ 「○」は班長

応募方法

応募自治体から直接、「応募様式」を電子メールで提出してください。

提出先： jichitai_data[at]e.u-tokyo.ac.jp

* “[at]”の部分を“@”に変えて送信してください。

提出期限：2023年6月16日（金）

- ※ 都道府県が取りまとめる必要はありません。
- ※ 応募について御関心がある場合は、未確定の場合も、提出期限より前に早めにお知らせいただけますと助かります。
- ※ 不明点、疑問点等あれば、遠慮なく問い合わせてください。
- ※ 期限に間に合わない場合は、期限までに御相談ください。
- ※ 応募の前に、HPにて公開しております「（参考資料1）個人情報保護制度との関係及び匿名化処理」「自治体データプロジェクトにおけるデータ取扱規則」「自治体データプロジェクトにおける安全管理措置等」もご一読ください。

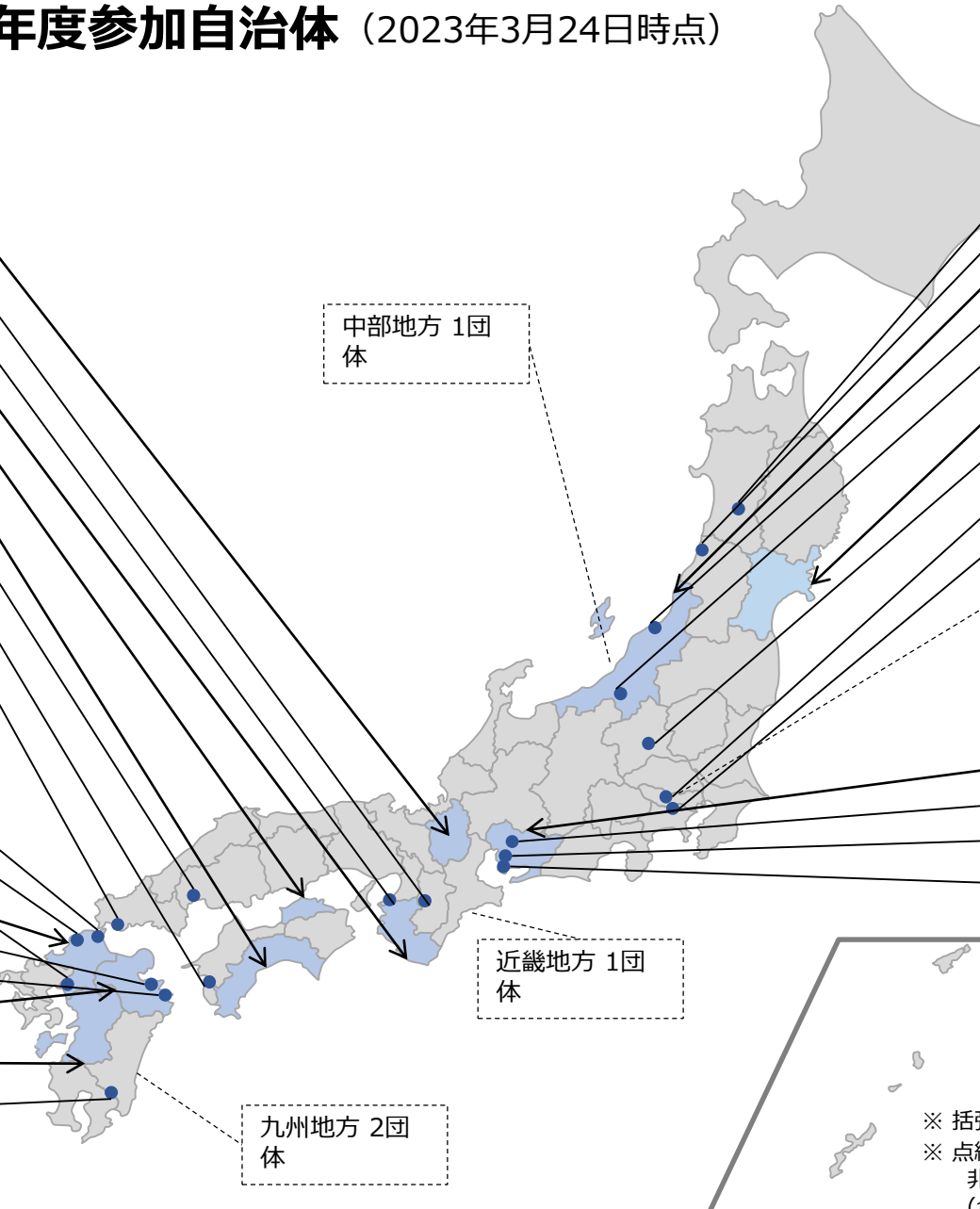
(参考) 2022年度参加自治体 (2023年3月24日時点)

滋賀県(1,413,610人)
五條市(27,927人)
阪南市(51,254人)
和歌山県(922,584人)
香川県(950,244人)
高知県(691,527人)
廿日市市(114,173人)
宇和島市(70,809人)
山陽小野田市(60,326人)

北九州市(939,029人)
宗像市(97,095人)
大川市(32,988人)
福岡県(5,135,214人)
大分市(475,614人)
津久見市(16,100人)
大分県(1,123,852人)
熊本県(1,738,301人)
都城市(160,640人)

大仙市(77,657人)
酒田市(100,273人)
新潟県(2,201,272人)
聖籠町(14,259人)
十日町市(49,820人)
宮城県(2,301,996人)
前橋市(332,149人)
羽村市(54,326人)
町田市(431,079人)

愛知県(7,542,415人)
長久手市(60,162人)
大府市(93,123人)
半田市(117,884人)



中部地方 1団体

関東地方 1団体

近畿地方 1団体

九州地方 2団体

※ 括弧内は令和2年国勢調査に基づく人口
 ※ 点線は自治体名の公表可否を確認中又は
 非公表を希望している5市
 (合計人口300~400万人)

Q&A (1)

全般

- Q1 匿名化に関する知識は必要ですか。
- A1 いいえ、必要ありません。個人情報保護及び匿名化の専門家の監修の下、デジタル社会形成整備法施行後の匿名加工情報制度において求められる水準の匿名加工の案を、CREPEから考え方とともに御提示します。
- Q2 匿名化等のため、プログラミングのスキルは必要ですか。
- A2 いいえ、必要ありません。匿名加工の案について自治体様の確認をいただいた後、CREPEが匿名化ツールを配布します。自治体様においては、CREPEが配布するマニュアルに従って匿名化ツールを実行いただくこととなります。
- Q3 データ分析のスキルは必要ですか。
- A3 いいえ、必要ありません。いただいた匿名化データをCREPEにおいて分析し、分析結果をお返しします。
- Q4 税務データと世帯データ、普通徴収のデータと特別徴収のデータなど、自治体側でデータ間の紐付けを行うことが必要ですか。
- A4 いいえ、必要ありません。紐付けのキーとなるデータ項目（宛名番号など）が入っていれば、CREPEにおいて紐付けします。

自治体側の負担

- Q5 参加に当たって、費用負担（CREPEへの委託料など）は必要ですか。
- A5 いいえ、必要ありません。ただし、システムからデータを抽出する際にシステムベンダに依頼しなければならぬ場合などに費用が発生する可能性があります。また、データを郵送していただく場合の郵送料や電磁的記録媒体等を準備していただく必要がございます。
- Q6 自治体側にどれくらいの作業負担が発生しますか。
- A6 2022年度に参加いただいた自治体の作業時間の中央値は15時間ですが、50時間を超える自治体も複数あり、自治体によって大きな差がありました。
- 2023年度は、2022年度に作業時間が長い自治体からアンケートを行い、自治体様の作業負担軽減に取り組みます。
- Q7 データ抽出作業にはどの程度の期間を要しますか。
- A7 自治体によってまちまちですが、ベンダへの依頼を行う場合、CREPEから自治体への匿名化ツール配布後、自治体様がデータを抽出し、CREPEに提供するまでに1か月～3か月を要する自治体がほとんどでした。自治体様の負担軽減に向けて当方も可能な限りの支援を行います。早めの御準備をいただくとスケジュールどおりの御提供ができるかと思えます。

Q&A (2)

匿名化

Q8 紐付け等のために、宛名番号・世帯番号を提供する必要がありますか。

A8 宛名番号・世帯番号そのものを提供いただく必要はありません。自治体様において設定いただくパスワードに基づいてハッシュ化された宛名番号・世帯番号を提供いただくこととなります。

ハッシュ化とは、元のデータから一定の計算手順に従ってハッシュ値と呼ばれる規則性のない値を求め、その値によって元のデータを置き換えることにより、データを不可逆的に別の形に変える方法をいいます。

例としては以下のようなイメージです。

宛名番号	ハッシュ化した宛名番号
12345678	aks;ldfjpawnefdoiewadlksfdajf
23456789	k3298refcsna3489elsjodslerire
34567890	kjlz-9erj8ufdjiawe8re;ofjwloe

ここで、「aks;ldfjpawnefdoiewadlksfdajf」から、元の「12345678」を復元することは不可能です。そのため、宛名番号について言えば、当方が宛名番号を知りえないまま、年度をまたぐ個人の紐付けが可能となります。

ハッシュ化については、CREPEが配布する匿名化ツールに組み込まれているため、自治体様においては、パスワードを設定いただければハッシュ化を行うことができます。

Q9 Rのバージョンについて、マニュアル記載のものと異なるバージョンでも問題なく動作しますか。

A9 Rは過去のバージョンとの互換性があるため、基本的に問題なく作業を行えます。

Q10 貸出PCでの匿名化作業となった場合、自治体負担の費用は発生しますか。また貸出PCはスタンドアロンでの運用が可能ですか。それとも自治体側のネットワークへの接続が必要でしょうか。

A10 貸出PCについて、費用負担はありません。また、貸出PCはインターネットには接続せず匿名化作業ができるよう、当センターで用意しております。

Q11 匿名化データやツールなど容量の大きなファイルの送受信について、どのような方法でやりとりしますか。

A11 オンライン上のデータ送信サービス、CD-Rの郵送など各自治体に都合の良い方法での送受信を予定しております。

Q&A (3)

データ内容・形式

Q12 現在、システムベンダごとにデータ構造・データ形式が異なっていますが、提供するデータのデータ構造・データ形式の指定はありますか。

A12 ありません。データ構造・データ形式の差異については、CREPE側で対応するので、心配いただく必要はありません。

具体的には、参加決定後、データ構造をお聞きし、それに基づいてCREPEにおいて自治体様のデータ構造に合わせた匿名化ツールを配布します。

Q13 何年分のデータが必要ですか。

A13 少なくとも5年分のデータは御提供いただいております。なお、機械学習を用いた予測を行うため、可能な限り多くの年度分のデータをご提供いただけますと予測精度の向上が期待できます。

Q14 ベンダーの変更に伴いいくつかの変更（①各種のマスクやデータベースの構成、データ項目が変更②新システム移行後は、データ提供時点のデータベース（現行システムのデータベース）からの抽出が制約される等）が想定されるが、プロジェクトへの応募は問題なく行えますか。

A14 変更後も応募要件を満たす場合は応募可能です。

Q15 収入と所得の変数が両方なくても参加可能でしょうか。

A15 給与、給与（専従）、雑所得（公的年金）以外については、所得金額があれば、収入金額がなくても応募可能とします。またシステムの管理上、給与、給与（専従）が分割されていない場合は給与のみで問題ありません。

Q16 自治体の全個人のデータベースがなくても参加できますか。

A16 大変申し訳ありませんが、自治体の全個人のデータベースをお持ちでない場合は、正確な分析ができないためご参加いただけません。

Q17 全てのデータベースに性別や生年月日がない場合でも参加可能でしょうか。

A17 全てのデータベースに年度と宛名番号があれば突合可能ですので、性別や生年月日が全てのデータベースに含まれている必要はありません。しかしながら、自治体の全住民が含まれるデータベースには性別、生年月日、続柄が含まれている必要がございます。

Q&A (4)

PCスペック等

- Q18 匿名化を行うPCのスペックはどの程度必要ですか。
- A18 自治体様の人口規模によって変動がありますが、昨年度ご参加いただいた人口10万程度の自治体様でプロセッサ：Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz 1.80 GHz、メモリ：16.0 GBのPCをお使いになった場合はデータ処理に大きな問題はなかったようです。
- 応募様式にご記入いただいたスペックを参考に、メモリ16GBの小型PC（モニタ・キーボード・マウスは自治体様に御用意いただきます。）を当センターの判断で貸し出す場合がございます。
- Q19 匿名化を行うPCはインターネットに接続できる必要がありますか。
- A19 インターネットに接続できる必要はありません。
- Q20 匿名化データの提出方法はどのようになりますか。
- A20 ファイル転送システム、記録媒体の郵送など、自治体様の都合の良い方法で結構です。送付方法の詳細については改めてご説明いたします。

- Q21 PCに外部のソフトウェアをインストールすることができません。匿名化にRを用いることは必須ですか。
- A21 匿名化等の作業に要する時間をできるだけ減らすため、必ずRで匿名化作業をしていただきます。

契約・協定

- Q22 契約や協定は必要ですか。
- A22 必ずしも必要ないと考えていますが、必要があると考える自治体は、応募様式でご回答ください。

分析

- Q23 最大何年間の推計が可能ですか。
A23 原則として1年間の推計が最大となります。これ以上の推計では予測精度の確保が難しいためです。
- Q24 今年度以降予定されている税制改正について、改正項目ごとに影響額を分析することは可能でしょうか。
A24 現状の税収予測モデルでは税制改正の影響を正確に推定することはできません。
- Q25 前年度の分析結果やサンプル、提供データを活用した研究は公開されているか
A25 前年度の分析結果やサンプルはCREPEのHP上で公表しております。また、提供データを活用した研究は完了次第、CREPEのHPで公表いたします。
- Q26 税収予測の精度はどれくらいですか。
A26 提供いただけるデータの数（人口）、年数等によりませんが、2022年度はほとんどの自治体で平均誤差率は1-2%でした（個人住民税）。
2023年度は、更なる精度向上を目指しています。