

行政記録情報等及び民間企業が保有する ビッグデータ等の活用

令和3年8月

総務省政策統括官（統計制度担当）

行政記録情報等の統計作成への 活用状況について

I 行政記録情報等の統計作成への活用状況

- 「公的統計の整備に関する基本的な計画」（令和2年6月2日閣議決定）においては、行政記録情報等の公的統計への活用の一環として、行政記録情報等の統計作成への活用実態を定期的に把握し、各府省に提供することが求められている。
- これを踏まえ、総務省政策統括官（統計制度担当）において、各府省等の協力を得て、行政記録情報等を用いて経常的に作成されている統計や、統計調査における行政記録情報等の活用状況の実態を取りまとめている。
- これらの結果は、総務省ホームページにて公表している。

https://www.soumu.go.jp/toukei_toukatsu/index/seido/seibi.htm

(注)「行政記録情報等」とは、国の行政機関が保有する各種行政記録情報（統計調査によって得られた情報を除く。）や地方公共団体が保有する業務記録情報を指す。具体的には、個別の法令の規定に基づいて為される申請、届出、登録、報告等によって得られる情報や、日々の業務活動（統計調査を除く。）を通じて収集・蓄積される情報が該当する。

I 行政記録情報等の統計作成への活用状況

(把握内容)

| I 行政記録情報等を用いて経常的に作成されている統計（業務統計）の概要 | II 行政記録情報等を活用している統計調査 |
|---|--|
| <ul style="list-style-type: none">(1) 政府統計コード(2) 統計の名称(3) 作成機関・部局課名(4) 作成目的(5) 作成開始年・年度(6) 作成周期(7) 公表方法 (公表月、HP掲載URL、e-Statへの掲載の有無、e-Statの掲載状況、その他)(8) 主な集計事項・作成指標(9) 業務統計の作成に活用されている行政記録情報等及びその収集等に関する根拠規定(10) 備考 | <ul style="list-style-type: none">(1) 行政記録情報等を活用している統計調査 (政府統計コード、統計調査名、調査実施機関)(2) 活用している行政記録情報等の概要 (行政記録情報等の名称、当該情報等の収集、作成又は整備に関する主な根拠規定、保有機関名、備考)(3) 統計調査への活用開始時期(4) 統計調査に活用する際の形態 (母集団情報の整備、調査事項の代替、欠測値補完、審査での活用等、具体的な活用方法)(5) 調査事項の代替による効果 (代替できた調査項目数、報告者数)(6) 活用条件、活用にあたっての制約・留意点 |

2 業務統計の作成状況等

(令和2年度統計法施行状況報告p141～)

○ 令和2年12月末現在で、行政記録情報等を用いて経常的に作成されている統計（業務統計）として各府省等から報告があったものは、以下のとおり

| 府省等 | 件数 | 例 | | 件数 | 例 |
|------|-------|-------------------------------|-------|--------|-----------------|
| 内閣官房 | 14(1) | 一般職国家公務員在職状況統計表 (人事統計報告) 等 | 財務省 | 37 | 財政統計 等 |
| 人事院 | 14(1) | 国家公務員給与等実態調査 等 | 文部科学省 | 31 | 公立学校施設実態調査 等 |
| 内閣府 | 6 | 児童手当事業年報 等 | 厚生労働省 | 88 | 医師・歯科医師・薬剤師統計 等 |
| 警察庁 | 2 | 犯罪統計 等 | 農林水産省 | 37 | 農業協同組合等現在数統計 等 |
| 消費者庁 | 9 | 地方消費者行政の現況調査 等 | 経済産業省 | 15(4) | 石油備蓄の現況 等 |
| 総務省 | 62 | 住民基本台帳人口移動報告 等 | 国土交通省 | 30 | 道路統計調査 等 |
| 法務省 | 15 | 出入国管理統計 等 | 環境省 | 26(4) | 騒音規制法施行状況調査 等 |
| 外務省 | 4 | 海外邦人援護統計 等 | 防衛省 | 13 | 自衛官の定員および現員 等 |
| | | | 計 | 398(5) | |

3 行政記録情報等を活用している統計調査

(令和2年度統計法施行状況報告p143～)

○ 統計調査の実施にあたり、母集団情報の整備や調査事項の代替など、行政記録情報等を活用している事例として各府省等から報告があったものは、以下のとおり

| | 統計調査数 | | | |
|-------|-------------|------------|-----------------|---|
| | うち 母集団情報の整備 | うち 調査事項の代替 | うち 欠測値補完、審査での活用 | |
| 内閣府 | 5 | 4 | 1 | 0 |
| 総務省 | 7(2) | 6(2) | 0 | 1 |
| 法務省 | 1 | 1 | 0 | 0 |
| 財務省 | 3 | 3 | 0 | 1 |
| 文部科学省 | 1 | 1 | 0 | 0 |
| 厚生労働省 | 27 | 16 | 12 | 3 |
| 農林水産省 | 11 | 8 | 6 | 0 |
| 経済産業省 | 11(2) | 11(2) | 2 | 0 |
| 国土交通省 | 33 | 31 | 13 | 0 |
| 環境省 | 5 | 4 | 1 | 0 |
| 計 | 102(2) | 83(2) | 35 | 5 |

(注1) ()内の数値は、共管統計(複数の府省が共同で作成する統計)の数であり、共管統計は、共管の府省にそれぞれ1件と計上しているため、各府省の対象統計数を単純合計しても、合計と一致しない。

(注2) 統計調査の件数は、統計法に基づく統計調査の承認単位で計上している。

(注3) 内数については、行政記録情報等の活用形態で1件と計上しているため、内数を単純合計しても、統計調査数と一致しない。

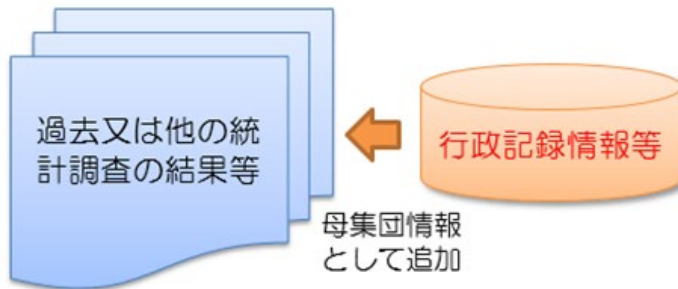
3 行政記録情報等を活用している統計調査

(令和2年度統計法施行状況報告p143~)

(参考) 統計調査における行政記録情報等の活用形態

【母集団情報の整備】

例1:



例2:

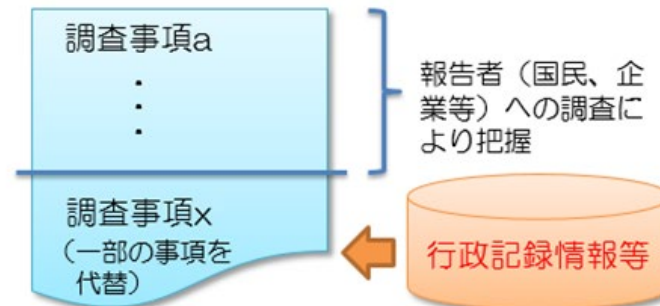


効果

的確な調査の実施、精度の確保・向上

【調査事項の代替】

例3:



効果

報告者の負担軽減、統計作成の効率化

3 行政記録情報等を活用している統計調査

(1) 母集団情報の整備

| 行政記録情報等の名称 | 統計調査名 | 根拠規定 | 具体的な活用方法 | 備考 |
|-----------------|--------------|--------------------|------------------------|--|
| 住民基本台帳 | 家計消費状況調査等 | 住民基本台帳法 | 調査対象の抽出、調査対象世帯名簿の作成 | <ul style="list-style-type: none">・閲覧手続や閲覧期間等が自治体によって異なるため、事務手続きが煩雑となるほか、閲覧に手数料を要する場合がある。・調査対象市区町村とは申請前に調整を実施 |
| 建設業許可データ、建設業者名簿 | 法人土地・建物基本調査等 | 建設業法 | 母集団情報の整備 | |
| 労働保険関係成立届 | 経済センサス-活動調査 | 労働保険の保険料の徴収等に関する法律 | 調査対象名簿の作成の際に労働保険データを活用 | <ul style="list-style-type: none">・同届は、事業単位で届け出ることとなっているため、「事業所」単位のデータを作成することが必要となる場合がある |
| 自動車登録ファイル等 | 自動車輸送統計調査等 | 道路運送車両法 | 自動車登録ファイルを用いて調査対象を抽出 | |

3 行政記録情報等を活用している統計調査

(2) 調査事項の代替

| 行政記録情報等の名称 | 統計調査名 | 根拠法令 | 具体的な活用方法 | 備考 |
|-------------------------|----------|-------|------------------------|---|
| 出生の届書等 | 人口動態調査 | 戸籍法等 | 出生等の届書の情報を基に調査票を作成 | ・戸籍法に基づく各届書に、人口動態調査に利用することを明記 |
| 医療施設の開設等の届出 | 医療施設調査 | 医療法等 | 届出情報を基に都道府県等において調査票を作成 | |
| 農業生産法人、認定農業者等に関する行政記録情報 | 農林業センサス | 農地法等 | 調査事項の一部代替 | |
| 建築着工等届出 | 建築着工統計調査 | 建築基準法 | 届出情報を基に調査票を作成 | ・建築基準法第15条第2項第4項において、都道府県知事は届出を基に建築統計を作成することが規定 |
| 輸出入申告情報 | 港湾調査 | 関税法等 | 調査事項の一部代替及び内容確認 | ・事前に報告者の同意を得るとともに、経費負担が必要 |

3 行政記録情報等を活用している統計調査

(3) 欠測値補完、審査での活用等

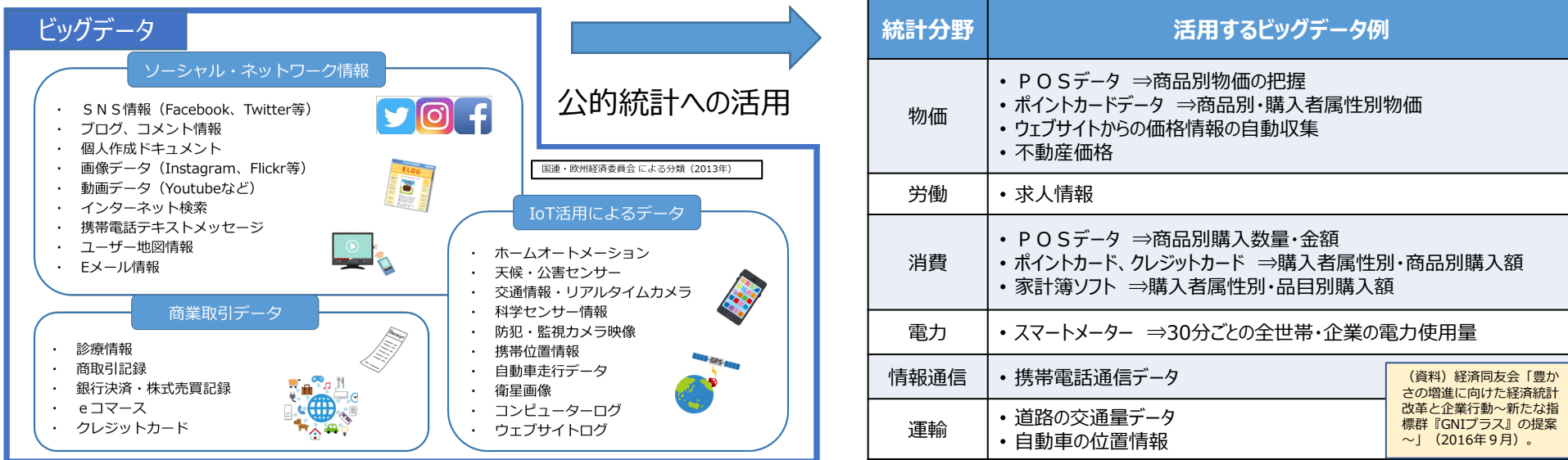
| 行政記録情報等の名称 | 統計調査名 | 根拠法令 | 具体的な活用方法 | 活用に当たっての留意点 |
|---------------|----------|-----------|----------------------|-----------------------------------|
| 住民基本台帳 | 国勢調査 | 住民基本台帳法 | 氏名等の調査事項に空欄があった場合に活用 | ・個人情報保護条例による手続等が必要となる場合がある |
| 有価証券報告書 | 法人企業統計調査 | 金融商品取引法 | 提出された調査票の数値チェック等に活用 | ・年次別調査において活用 (四半期報告書は連結情報のための) |
| 雇用保険適用事業所設置届等 | 毎月勤労統計調査 | 雇用保険法施行規則 | 母集団労働者数の補正に活用 | |

ビッグデータ等の活用に関 係する検討状況

「ビッグデータ等の利活用推進に関する産官学協議のための連携会議」の概要

趣旨

「公的統計の整備に関する基本的な計画」（平成30年3月6日閣議決定）を踏まえ、ビッグデータ（以下「BD」という。）等の活用に係る先行事例があるデータ又は優先度の高いデータ等（行政記録情報等を含む。以下同じ。）を選定して、関係者の協力を得て集中的に協議することにより、利活用上の各種課題の解決や優良事例等を積み上げて公表するなどし、各府省、地方公共団体、民間企業等におけるデータ等の相互利活用を推進する。



構成員

- 高橋 大志 慶應義塾大学大学院教授 (座長)
 庄司 昌彦 武蔵大学教授
 田原 健吾 日本経済研究センター主任研究員
 水野 貴之 国立情報学研究所准教授
- (民間)
 大西 浩史 日本データマネジメント・コンソーシアム (JDMC) 理事 兼 事務局長

開催実績

- 第1回 (H30.5.23) BDの利活用に向けて
- 第2回 (H30.6.27) 労働市場の転職時賃金情報の活用 (R社)
- 第3回 (H30.7.25) 携帯GPSを活用した流動人口把握 (A社)
- 第4回 (H30.10.4) 民間企業のエリアマーケティング活動におけるBD活用 (Z社)
- 第5回 (H30.12.5) PT調査でのBD活用 (国交省)、CTIでのBD活用 (統計局)
- 第6回 (R1.5.22) 流動人口検証WG報告、商業動態統計試験調査 (経産省)
- 第7回 (R1.7.31) 電力データ活用の取組み (G組合・A区)、労働BD活用 (R社)
- 第8回 (R1.11.13) CPIへのウェブスクレイピングの活用 (総務省)
- 第9回 (R2.2.26) 自治体都市開発におけるビッグデータ活用の取組 (K市・Y社)
- 第10回 (R2.9.30) 地球観測データを活用したSDGsの算出 (JAXA)、活用事例整理
- 第11回 (R2.12.23) 観測データWG報告、労働BD利活用 (R社)
- 第12回 (R3.03.10) 位置情報による経済活動のナウキャスト (N銀)、観測データWG報告
- 第13回 (R3.06.30) データマネジメントの重要性 (JDMC)、損害保険に関するBDの活用 (MS社)

目的

「ビッグデータ等の利活用推進に関する産官学協議のための連携会議」で紹介された、ビッグデータの利活用における先進事例を通じて明らかとなった、ビッグデータの活用法やメリット、課題とその解決事例、及び今後対応していく方向性など、関係者において共有することが有用である内容を整理し、公的統計へのビッグデータ利活用の促進を図る。

(1) 既存統計の補完

ビッグデータ活用のメリット

活用事例 (*は研究段階のもの)

(1) - ①. 速報性

速報性を有するオルタナティブデータを活用することにより、社会経済の変動を適時に把握し、必要な対策をより早期にとることが可能となるほか、市場へのインパクトが分散されること等も期待。

総務省が参考公表を行っている消費動向指数 (CTI) *では、POSデータ等を活用し、速報性に優れた指標を開発中。

(1) - ②. 詳細化

ビッグデータを活用して公的統計の内訳を明示することにより、より詳細な分析に基づいた対策の検討や、既存統計が社会環境の変化へ適用できているかの検証を行うことができる。

国土交通省が平成30年に実施した東京都市圏パーソントリップ調査では、調査結果から得られた地域別のトリップ数 (移動した回数) をベンチマークとし、携帯基地局情報から作成した内訳比率を活用し、より詳細な地域単位でのトリップ数の推計を試みている。

(1) - ③. カバレッジ拡大

ネット販売の拡大により調査票による従来手法の統計作成では補足できないデータが増加しているが、網羅的にウェブデータを収集することによりデータのカバレッジを拡大し、統計精度を向上させることが期待。

総務省の消費者物価指数 (CPI) では、ネット販売が主流の「外国パック旅行費」「航空運賃」「宿泊料」について、ウェブスクレイピング (ウェブ上の情報を自動的に収集・保存する技術) による高頻度かつ大量の価格収集を行い、統計の精度向上を進めている。

(1) - ④. 新指標作成

統計を補完するためには、直接的なデータのみでなく全く別の視点から抽出したデータによる擬似的な指標の作成や、複数のデータを組み合わせて新たな統計を作成することも有効であると考えられ、試行や研究が進められている。

経済産業省では、スーパーマーケット、コンビニエンスストア、ホームセンター、ドラッグストア等の小売各業態のPOSデータの活用により、消費者心理を捉える指標 (METI×NOMURAコンシューマセンチメントインデックス*) のほか、週次×地域×商品別の販売動向を把握することが可能な指標 (METI POS小売販売額指標 [ミクロ] *) の開発を行い、「BigData-STATSダッシュボード (β版)」で公表*を行っている。
※ METI×NOMURAコンシューマセンチメントインデックスは2021年6月末をもって公開終了。

(2) 報告者負担の軽減

統計調査においては、調査員の不足や報告者の負担軽減などへの対応とともに、データを収集・集計する行政側での業務負担の軽減が求められており、ビッグデータの活用はこれを解決する有力な手段としても期待。

総務省の消費者物価指数 (CPI) におけるウェブスクレイピングによる自動的な価格収集や、経済産業省の商業動態統計調査におけるPOSデータの活用による家電大型専門店に関する調査の代替により、報告者負担・業務負担の軽減を実現。

公的統計へのビッグデータの更なる活用に向けて_BD連携会議での事例整理②

| ビッグデータ活用における課題 | 解決方法・取組の方向性 |
|------------------------------|---|
| (1)ビッグデータが有する偏り等への対応 | <p>○ビッグデータの公的統計への利活用に当たっては、データの偏りの把握、精度の検証、公的統計との整合性を勘案したデータ選定など、ビッグデータ自体の特性を十分に把握した上で検討を進めることが必要。</p> <p>○データ提供者側からデータの特徴（データソース、作成プロセス、提供データの精度や偏り、利活用実績等）を開示していただけるよう、行政側としてチェックリストを整備し、その活用を働きかけていく取組も必要。</p> |
| (2)安定的・継続的な民間データの提供 | |
| (2)－①データ取扱いのルール設定のための標準規約の整備 | <p>○信頼関係の構築には、オフィシャルな関係作りとともに、信頼関係に基づく担当者間での密な連携が欠かせない。目的・対象データ・使用期間などの標準項目を設定した標準的な規約のひな形を整備し、締結の実績を蓄積していくことは有効。</p> |
| (2)－②データ提供者へのインセンティブの付与 | <p>○提供者へのインセンティブとして、データ提供の対価を支払うことが挙げられるが、研究機関では、対価としてデータ分析結果や研究レポートの発行など、データの品質を保証することで無償提供を受けている事例もある。提供者側のインセンティブも考慮しつつ、官民でWin-Winの関係を構築する必要がある。</p> |
| (2)－③データ提供が中断するリスクへの対応 | <p>○企業の経営方針の変更や倒産などにより、継続的なデータ提供を受けられなくなる可能性がある。また、データ購入に当たり、購入価格の上昇や入札による提供元の変更などにより、統計の継続性の確保が困難になるリスクがある。事態発生時の対応を事前に検討しておく必要がある。</p> |
| (3)人材育成と体制整備 | <p>○会議で紹介された事例では、業務の品質を確保するため、組織的な人材育成・確保やノウハウの蓄積・継承が行われているが、民間との人材交流を行うことにより、官民の信頼関係の醸成や人脈作りにつながることを期待される。民間人材の行政における活用のみならず、行政職員の研究機関等への出向も有効。</p> |
| (4)システム・技術の整備 | <p>○会議で紹介された事例はデータ処理に特化したものであったが、DXの時流に沿って全体最適のシステムとするためには、プロセス全体を見据えたトータルなシステム構築の観点も重要である。また、ウェブ掲載データ・フォーマット等の標準化を進めることで、社会全体として効率的なデータエコシステムの構築につながる。</p> |
| (5)その他（情報保護、法令関係など） | <p>○民間企業が安心してデータを提供できるよう、データの流通環境を整備する必要がある。また、個人情報については、データ提供者による担保のほか、その取扱いに関する社会的なコンセンサスの確立を図る必要がある。</p> <p>○ビッグデータの利活用環境や技術の発展・変化に合わせ、公的制度におけるビッグデータの位置付けの整理を進めることも重要。</p> |

総務省におけるビッグデータの活用に係る取組

総務省政策統括官（統計制度担当）では、「ビッグデータ等の利活用推進に関する産官学協議のための連携会議」における優良事例の共有等の取組に加え、現在、ビッグデータの活用に係る以下の取組を実施している。

- ① **流動人口データを活用した統計調査の補完可能性の検討**（平成30年度～）
メッシュ型流動人口データを活用し、観光関連の統計調査との比較シミュレーションを実施。
- ② **労働ビッグデータの統計的利活用に関する研究**（令和元年度～）
労働ビッグデータを活用し、転職時の賃金変動状況の予測について研究。
- ③ **地球観測衛星データを活用したSDG指標に関する検証**（令和2年度～）
地球観測衛星データを活用し、SDG指標15.4.2（山地グリーンカバー指数）及び11.3.1（人口増加率と土地利用率の比率）について試算・検証。
- ④ **ビッグデータ等を活用した統計作成の推進に向けた調査研究**（令和3年度）
各国の統計機構におけるビッグデータ等の利活用事例を網羅的に整理。また、新型コロナウイルス感染症の拡大を契機に活用されるようになったデータや事例の洗い出し等を実施。

（※）なお、これまでの取組も踏まえ、令和4年度以降、ビッグデータ等の活用事例に関する情報を一元的に集約・提供し、ビッグデータ等の利活用に携わる関係者を結びつけるプラットフォームとなる「ビッグデータ・ポータル（仮称）」を構築する方向で検討を進めている。

ビッグデータの公的統計への利活用事例等(詳細)について①

| 分野 | 物価 | 消費 | 商業 | 交通 |
|----------|---|---|--|--|
| 公的統計 | 消費者物価指数(CPI) | 消費動向指数(CTI) | 商業動態統計(家電大型専門店分野) | パーソントリップ(PT)調査 |
| ビッグデータ | WEB掲載価格データ | POSデータ・クレジットカード等 | POSデータ | 携帯基地局情報 |
| 統計の目的 | 全国の世帯が購入する家計に係る財及びサービスの価格等を総合した物価の変動を時系列的に測定する。家計の消費構造を一定のものに固定し、これに要する費用が物価の変動によって、どう変化するかを指数値で示したもので、毎月作成。指数計算に採用している各品目のウエイトは家計調査の結果等に基づいている。毎月の品目価格は小売物価統計調査によって得られる。 | 家計調査の結果を補完し、消費全般の動向を捉える分析用のデータとして開発中の参考指標。家計消費指数を吸収するとともに、単身世帯を含む当月の世帯の平均的な消費(CTIミクロ)、家計最終消費支出の総額の動向(CTIマクロ)を推計している。 | 全国の商業を営む事業所及び企業の販売活動などの動向を明らかにすることを目的として実施。全国の卸売・小売事業所を対象として、商品販売額、販売先別商品販売額、月末従業員数、期末商品手持額等を調査している。 | 都市における人の移動に着目した調査。世帯や個人属性に関する情報と1日の移動をセットで尋ねることで、「どのような人が、どのような目的で、どこからどこへ、どのような時間帯に、どのような交通手段で」移動しているかを把握することが可能で、都市交通の現況の把握、将来交通需要の予測、都市交通マスタープランの作成等、都市交通に関する計画等を策定する上での基礎資料とすることを目的としている。 |
| データの収集方法 | WEBにおける公表情報をスクレイピングにより収集。 | 消費動向指数研究協議会を通じて、POSデータ保有企業、クレジットカード会社等から提供。 | POSデータを収集・分析するマーケティングリサーチ(MR)会社における加工を経て、情報を収集。 | 携帯キャリア会社から提供。 |
| 概況【背景】 | 【背景】 統計改革の基本方針等において、インターネット販売価格の更なる捕捉に向けた検討を行うとの提言がなされたことを受け、CPIにおけるインターネット価格のより精緻な把握に向けた検討を開始。 | 【背景】 消費全体の動向を、マクロ、ミクロの両面でもとらえる、速報性を備えた包括的な消費関連指標の在り方について検討することを目的として、総務大臣主催の「速報性のある包括的な消費関連指標の在り方に関する研究会」を開催し、消費動向指数を開発。 | 【背景】 BDを活用した新指標の開発に向けて調査研究を進める中で、公的統計へのBD活用が統計改革の基本方針、第Ⅲ期基本計画に明記されたことも踏まえ、BDを活用した商業動態統計の実施に向け検討を開始。 | 【背景】 1970年代以降、PT調査の実施とこれに基づく総合都市交通計画の立案が継続的に行われて、科学的な分析結果に裏打ちされた都市交通施策の推進に大きく寄与してきた。一方で、①近年の自治体における都市交通上の課題として「短中期的・ミクロな交通政策」までニーズが拡大したこと、②全国あらゆる場所で、24時間365日データが蓄積される交通系ビッグデータが登場したこと、など状況が変化してきたことを受け、PT調査へのビッグデータ活用の検討を開始。 |
| 概況【現状】 | 【現状】 有識者を交えた「物価指数研究会」の検討、平成30年度統計法施行状況に関する審議を経て、旅行サービス3品目について、消費者物価指数2020年基準の2020年1月分からウェブスクレイピングの実運用を開始。現行の2015年基準は2021年12月分公表迄であり、現行と並行して進めていく予定。 | 【概況】 消費動向指数研究協議会を設立後、消費動向指数の研究を継続的に実施し、既存統計をデータソースとした消費動向指数(CTI)を開発し、平成30年1月分から参考指標として公表開始。今後、研究分析・検証を得た後、ビッグデータを順次活用する予定。 | 【現状】 2018年に試験調査を実施し一定の目処が立ったため、2020年より本格的に実施。BDを活用するかどうかは企業の意向次第(調査票記入とPOSデータの活用の選択制)。 | 【現状】 PT調査データとビッグデータを組み合わせ、それぞれの強みを生かすことのできる総合交通調査体系の構築を目指す一環として、「総合都市交通体系調査におけるビッグデータ活用の手引き」を作成。H30年の東京都市圏PT調査でも本手引きを活用し、調査設計・分析等を実施。今後、近畿圏PT調査・中京圏PT調査でも活用していく予定。 |

ビッグデータの公的統計への利活用事例等(詳細)について②

| 分野 | 物価 | 消費 | 商業 | 交通 |
|---|--|--|---|---|
| 公的統計 | 消費者物価指数(CPI) | 消費動向指数(CTI) | 商業動態統計(家電大型専門店分野) | パーソントリップ(PT)調査 |
| ビッグデータ | WEB掲載価格データ | POSデータ・クレジットカード等 | POSデータ | 携帯基地局情報 |
| メリット | | | | |
| 主目的 | Q WEBスクレイピング技術で大量のネット価格(旅行関係)を取得することによる、CPIの精度向上。 | Q POSデータ等を活用した高精度で消費動向を捉えることのできる速報性の高い指標の構築。 D (当面は、CTIマクロにおけるGDP統計の構成要素である家計消費支出の予測精度向上の実現を目指して検討中。) | B 家電大型専門店に関する調査をPOSデータで代替すること(MR会社によるPOSデータの調査票情報への変換)による、報告者負担の軽減。 | Q トリップ総量をPT調査、各ゾーンにおけるトリップ目的・トリップ手段に関する比率を携帯基地局データで算出することによる、目的別手段別小ゾーン間ODの推計。 |
| 副次効果 B: 報告者負担の軽減 Q: 統計の精度向上、詳細化 C: 統計作成の効率化、コスト削減 D: 統計公表の早期化 | B 価格データの整理・提出のために実施されるヒアリングが簡素化される。(webスクレイピング以前は、調査対象企業のHPを自らチェックするだけでなく、企業ヒアリングもしており、企業側でそのヒアリングに対応する作業が少なくなる。) | Q 民間企業が保有する様々な消費関連情報を活用した消費動向指数の開発について産学官で連携して研究を行い、企業保有情報の適切かつ有効・有益な活用により、我が国の公的統計の改善・高度化、学術研究の発展を推進。 | Q POSデータを統一的に集計することで、それぞれの記入者が異なった分類に記入するといった非標本誤差の縮減につながる。 | B ゾーン単位での適切な目標精度の見極めによるサンプル率の減少(2%→1%)に伴う、報告者数の最適化。 |
| | C これまでの目視作業による手法(特定日・時間に職員等が自らHPの価格情報をチェック)から、WEBスクレイピングによる自動的な価格収集になることで、職員による作業が軽減される。 *一方で、収集データの分析作業、データクレンジング、膨大なデータの保管方法等新たなコストも発生。 | | | C ゾーン単位での適切な目標精度を確保した統計作成の効率化・コスト削減。 |
| 課題・取組 | | | | |
| ビッグデータの選定・加工 | 【データ選定と対象分野の絞り込み】 品質のバラツキを少なくできるため、WEBスクレイピングによる価格収集に馴染む、航空運賃・宿泊費・外国パック旅行を対象とした。 | 【データ選定と対象分野の絞り込み】 CTIマクロの検討に当たって、POSデータ、クレジットカードデータ等を候補として検討中。 | 【データ選定と対象分野の絞り込み】 商業動態統計の対象業種のうち、商品・店舗(企業のみならず事業所も含む)に関して網羅的なデータが既に把握されており、かつ当該データの活用がビジネスとして確立している、家電大型専門店分野を対象とした。 | 【データ選定と対象分野の絞り込み】 交通系ICカードから得られるビッグデータなどもあるが、交通手段やエリア的な網羅性が最も高い携帯基地局情報を選択した。調達先は安価で協力してくれることを重視して選定した。 |
| | 【処理方法】 各品目別に取得データの拡大に対応した処理方法の決定。 (データの拡大) ・取得間隔: 毎日 ・パラメータ種類(航空運賃): 経路・時間帯・割引プラン (品目別) ・航空運賃は品質の違いは少ない ・宿泊費は品質がバラバラのため、観光統計から宿泊キャパシティが大きい宿を選定 ・外国パック旅行は旅行会社毎に品質差異があるが、エアライン・ホテルの水準を合わせて選定 | | | 【データ特性把握】 以下のようなデータ特性を踏まえて、データ利用を行う必要がある。 ・基地局密度の小さいエリアや人口の少ないエリアのデータではデータ数が少なく精度に懸念があったり、個人が特定される恐れがありデータ提供がされない可能性 ・小学生低学年は携帯電話の保有率の低さによるデータ精度の懸念及びデータ入手ができない可能性 ・滞在時間によっては、経路地と目的地の区分けが正確にできない可能性(定義の再検討が必要) |

ビッグデータの公的統計への利活用事例等(詳細)について③

| 分野 | 物価 | 消費 | 商業 | 交通 |
|-----------|---|--|---|--|
| 公的統計 | 消費者物価指数 (CPI) | 消費動向指数 (CTI) | 商業動態統計 (家電大型専門店分野) | パーソントリップ(PT)調査 |
| ビッグデータ | WEB掲載価格データ | POSデータ・クレジットカード等 | POSデータ | 携帯基地局情報 |
| 課題・取組(続き) | <p>【信頼関係】 サイト運営者に対しては、丁寧に説明し協力要請。 スクレイピングはサーバーに負荷をかけるため、データ取得の時間帯・頻度・IPアドレスなどを事前に伝達し、許可を得たうえで、サイト運営者から承認を得た方法で実施している。</p> <p>【契約】 データ利用に関し、目的外利用禁止を誓約している。</p> | <p>【信頼関係】 消費動向指数研究協議会の開発に賛同してもらっている企業に、消費動向指数研究協議会に参画してもらっている。同企業に対しては、訪問面談を通じて、最低年1回はデータホルダーと接点を持ち、データ提供者との信頼関係を構築している。</p> | <p>【信頼関係】 報告者(小売企業)に対しては、経産省として足を運び、POSデータの活用による調査方式への協力を依頼している。 *従来調査では、傘下グループ企業がある場合、ホールディング企業が複数企業分をまとめて調査票記入していた例もあり、企業の負担が大きかったが、POSデータ活用によるそれらの作業が軽減するため、協力を受けやすかった。</p> | |
| | <p>【継続性】 WEBスクレイピングは高い技術力がなくても実施できるため、対応業者は複数ある。また業者変更になってもデータの継続性は容易に確保できると考えている。</p> | <p>【継続性】 参画企業に対し、毎月の公表結果をメールにて報告している他、研究の進捗などを定期的に提供。 現在、データの継続的な提供に向けて調整中。</p> | | <p>【継続性】 ビックデータ活用に関する作業をすべて委託研究で実施。 委託研究は競争入札となり、入手されるデータの信頼性や精度にバラツキがでる可能性がある。 (そのため、総数把握には使用せず、計画基本ゾーン等内の比率設定に使用するなど、当該データの特性を前提とした活用方法としている。)</p> |
| | <p>【ビジネスモデル】 WEBスクレイピングは(サイト運営者のサーバーに負荷はかけるが)公表情報を収集するものであるため、データの取得そのものの購入費等は不要。</p> | <p>【ビジネスモデル】 現在は、データホルダー(個別企業等からデータを収集している企業)から、研究のために不定期に無償でデータ提供を受けている。</p> | <p>【ビジネスモデル】 MR会社における家電量販店POSデータの収集・分析業務は、従来より家電メーカーなどを顧客としたビジネスモデルが確立されている。 今回はMR会社に対して組替え集計業務を委託するものである。</p> | <p>【ビジネスモデル】 調査委託先でのBDの調達コストが上昇する可能性が高く、今後も適切な予算を確保することが課題となる。</p> |

ビッグデータの公的統計への利活用事例等(詳細)について④

| 分野 | 物価 | 消費 | 商業 | 交通 |
|---------------|---|--|---|--|
| 公的統計 | 消費者物価指数 (CPI) | 消費動向指数 (CTI) | 商業動態統計 (家電大型専門店分野) | パーソントリップ(PT)調査 |
| ビッグデータ | WEB掲載価格データ | POSデータ・クレジットカード等 | POSデータ | 携帯基地局情報 |
| 課題・取組 (続き) | 法規制・契約(外的制約) | 【個人情報保護】 個人情報保護に関して、協議会で規約を定めて、それに従って参画企業からデータを受領している。 | 【個別契約】 もともと小売企業とMR企業の間でデータ授受の契約をしており、その中に経産省への組替え集計結果の提供を可能とする覚書を追加した。 | 【個人情報保護】 データホルダーは個人情報保護法に則り、必要に応じて利用者に許諾を取る必要がある。 |
| | 【統計法】 今回のようなデータの収集方法と、統計法上の扱いについて整理中。 (HP上の公表情報を収集する作業であり、「統計調査」には該当しないものと想定。) ※2020年消費者物価指数の基準改定を踏まえ、小売物価統計調査の調査計画の変更について、統計委員会に諮っている(2020年6月)。 | | 【統計法】 統計法上の「統計調査」と今回のようなデータ収集方法がどのような関係にあるのか、整理に時間を要した。 | |
| 課題・取組 (続き) | 運用 | 【コンピューターシステム】 特別なシステムは構築していないが、品目毎に価格設定ロジックを内製で作り込んでいる。 | 【コンピューターシステム】 ビッグデータをデータソースとした消費動向指数の開発をセキュアかつ効率的に行うことを目的としたクラウド環境を構築し、運用。データ提供元の利用環境による環境依存性の排除が課題。 | 【コンピューターシステム】 特別なシステムは構築していない。 *これまでと同項目の調査票情報を仲介するMR会社に集計・提出してもらうものであり、省内のインターフェースを変更する必要はない。 |
| | 【運用体制】 データ収集等作業は業務委託。 (これにより、土日祝日における収集作業等も可能。) | 【データ提供】 特定のデータだけに依存しない方針の検討 | 【データ公開時期】 POSデータの活用によりデータの収集は早められる可能性があるが、提供先企業における財務情報開示のタイミング等の関係から、必ずしも早期に公開できるとは限らない。 | 【コンピューターシステム】 特別なシステムは構築していない。 |
| | 【データ管理】 データ量が膨大(数十GB/月)なので、統計の基礎データとして管理するには相応の容量のストレージを必要としており、これを継続していくためには予算化が必要。 | | | |
| | 【ノウハウ継承】 スキル・ノウハウが検討担当者に偏在しつつあるため、データサイエンティスト育成の取組とセットでスキル継承を行う取組が必要である。 | | | 【人材確保】 データサイエンティスト人材の確保の予定はなく、通例として必要なときに外部委託している。手引き書を作成することにより近畿圏・中京圏等他の主体への展開やスキル継承支援を可能としている。 |

ビッグデータの公的統計への利活用事例等(詳細)について⑤

| 分野 | 物価 | 消費 | 商業 | 交通 |
|------------------|--|--|------------------------------------|---|
| 公的統計 | 消費者物価指数 (CPI) | 消費動向指数 (CTI) | 商業動態統計 (家電大型専門店分野) | パーソントリップ(PT)調査 |
| ビッグデータ | WEB掲載価格データ | POSデータ・クレジットカード等 | POSデータ | 携帯基地局情報 |
| その他 | | | | |
| 検討の仕組み (URL) | 物価指数研究会 https://www.stat.go.jp/info/kenkyu/cpi/index.html | 消費動向指数研究協議会 | ビッグデータを活用した新指標開発に関する研究会 | 携帯電話基地局の運用データに基づく人の移動に関する統計情報の交通計画等への適用に関する共同研究 |
| ドキュメント等 (URL) | 統計委員会(平成30年度統計法施行状況に関する審議結果報告書) https://www.soumu.go.jp/main_sosiki/singi/toukei/kikaku/kaigi/02shingi05_02000367.html | 協議会規約(個人情報保護) http://www.stat.go.jp/data/cti/pdf/ho20171226.pdf | | 総合都市交通体系調査におけるビッグデータ活用の手引き https://www.mlit.go.jp/common/001241230.pdf |
| 所管組織 | 総務省統計局統計調査部消費統計課物価統計室 | 総務省統計局消費統計課 | 経済産業省大臣官房調査統計グループ調査分析支援室/サービス動態統計室 | 国土交通省都市局都市計画課都市計画調査室 |
| BD連携会議* | 第8回 | 第5回 | 第6回 | 第5回 |

(*)

https://www.soumu.go.jp/main_sosiki/kenkyu/big_data/index.html