

公的統計へのビッグデータの更なる活用に向けて

－ビッグデータ連携会議におけるこれまでの事例整理－

ビッグデータ連携会議決定
令和2年12月23日

1 本資料の目的

「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¹」（本資料において「ビッグデータ連携会議」という。）は、「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、統計的分析や統計作成におけるビッグデータ等の活用に係る事例を通じ、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ことにより、各府省・地方公共団体・民間企業等でのデータ等の相互利活用を推進することを目的として開催している。

本資料は、第9回までの会議において紹介された、民間の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（当資料では○で記載）、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ことを目的としている。

2 ビッグデータの活用法やメリット

社会におけるDX（デジタルトランスフォーメーション）の取組が進められ、また、様々な場面でのビッグデータの利用が広がる中で、統計分野においても、統計調査における報告者の負担軽減のみならず、ニーズに即した統計の作成等にも寄与することが期待されている。

(1) 既存統計の補完

ビッグデータは公的統計の作成を目的として収集されたものではなく、データそのものや収集した組織や環境に起因する偏りを持っているので、既存の統計を完全に代替するためには、現時点では多くの課題を解決しなければならない。一方、その特徴である高頻度かつ多種大量にデータが生成されることを活かし、既存統計では把握できない時間的・空間的・統計分類的観点等からの既存統計の補完や、新たな指標の作成などにより、公的統計をより利活用性の高いものとするのが可能である。

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される。月例経済報告に関しては、クレジットカードの決済情報を使用した消費動向を捉える新しい指標がオルタナティブデータとして紹介されている。また、消費動向指数（以降、CTI²）では、公表の早期化などを念頭に、POSデータ等の民間ビッグデータを活用し、従来

¹ ビッグデータ連携会議：https://www.soumu.go.jp/main_sosiki/kenkyu/big_data/index.html

² 消費動向指数：<https://www.stat.go.jp/data/cti/index.html>

よりも速報性に優れた指標を開発中である。

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。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、統計調査を「総量」のベンチマークとして、その「内訳比率」をビッグデータから作成し、ゾーン内部OD（出発地－目的地）を統計調査よりも詳細に分割した目的手段別OD表（起終点表）を推計している。

③ カバレッジの拡大

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できる。消費者物価指数（以降、CPI³）では、ネット販売が主流となっている「外国パック旅行費」「国内航空運賃」「国内宿泊料」に対して、ウェブスクレイピングにより旅行会社・プラン数や路線数の把握対象の拡大、価格収集期間や高頻度（毎日）の価格把握などを行い、統計の精度向上を進めている。

④ 新たな指標等の作成

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。経産省では、民間事業者などと連携し、Twitterから景気に関連するキーワードを含むツイートを抽出し、センチメント評価を行うことで景況感を表す指数⁴などを始め、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、ドラッグストアによるPOSデータから、消費者心理をとらえる指標⁵などの開発を試みている。

(2) 報告者負担や業務負担の軽減

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されている。CPIではウェブスクレイピングにより、該当費目の報告者負担をほぼゼロにするとともに、行政側における効率的なデータ収集を実現している。また商業動態統計⁶では、家電大型専門店に関する調査をPOSデータでの代替も可能とすることにより報告者負担の軽減を実現している。

³ 消費者物価指数：<https://www.stat.go.jp/info/kenkyu/cpi/index.html>

⁴ SNS×AI 景況感指数、鉱工業生産予測指数：http://qr.nomura.co.jp/quants/sns_ai/

⁵ METI×NOMURA コンシューマーセンチメント・インデックス（消費者心理指標）：https://www.meti.go.jp/statistics/bigdata-statistics/bigdata_pj_2019/pos_nomura.html

⁶ 商業動態統計：<https://www.meti.go.jp/statistics/tyo/syoudou/index.html>

3 ビッグデータの活用における課題及びその解決方法

ビッグデータの活用を推進するためには、データの精度検証やデータの入手方法の確立、人材面等の課題が存在し、これらの解決を図ることが必須である。

(1) ビッグデータの特徴把握

ビッグデータはデータ生成等に起因する偏りや特徴を有しているため、統計への利活用にあたっては、データの偏りの把握、精度の検証、及び公的統計との整合性を見極めてデータ選定するなど、ビッグデータ自体の特徴を十分に把握した上で検討を進めることが必要である。

- 東京都においてGPS方式のメッシュ型流動人口データの有効性の検証を行い、アプリ利用者属性に起因する標本の偏りが存在することを確認した。その上で、国勢調査や基地局データと相関を有することや、大量の人の動線のハブとなる都心ターミナル駅や海岸沿い等隣接するメッシュ間の人口差が大きいエリアで比較的有效性に機能しうることを確認した。(メッシュ型流動人口検証WG⁷)
- 一般にウェブスクレイピング情報は店頭・ネット販売価格全体を代表していないことに留意し、ネット販売が主流であることやデータ特性等の費用対効果を見極め、価格動向をウェブスクレイピングにより把握する品目として、「外国パック旅行費」「国内航空運賃」「国内宿泊料」を選定した。例えば「外国パック旅行費」については、旅行会社から得られた販売実態に関する詳細な情報を元に、膨大な数の商品情報からノイズを除去し、物価動向把握において前提となる同品質の商品価格の追跡が可能となった。(CPI)
- POSデータの活用にあたり、一定の店舗数を保ちながらデータを活用する必要があるため、仮にこれらの店舗の中から閉店になるものがあつた場合でも、同規模の店舗を割り当てることにより、データの断層を極力少なくするなどの取組が検討されている。(CTI)

上記のように、統計を作成する側においてデータの特徴を把握し、精度や偏りを検証する取組が必要であるものの、そうした取組を実効性のあるものとするためには、データホルダーがビッグデータを提供する際に、そのデータが有する特徴(例:データソース、作成プロセス、提供データの精度や偏り、データの完成度、利活用実績)を開示することが有効であると考えられる。行政側としてはこうしたデータの特徴を確認するためのチェックリストを整備し、データホルダーに活用してもらうよう働きかけていく取組が必要である。

(2) ビッグデータの入手

民間からビッグデータを継続的に提供してもらうためには、官民の信頼関係を基礎にして、民間へのインセンティブ付与やデータ提供の中断に対するリスク対応などにも配慮した取組を進める必要がある

⁷ WGのワーキングペーパー「東京都における流動人口データの有効性の検証」

https://www.soumu.go.jp/main_content/000630006.pdf

① データ取扱いのルール設定

民間から提供されたデータの取扱いに関しては、協定書の締結に基づくオフィシャルなルール設定や、そのルールに基づいて窓口担当者間での実務的な交流を行うなど、データ提供を継続してもらうために必要となる信頼関係を構築・継続するための取組が行われている。

- CTIの取組では産学官の連携による「消費動向指数研究協議会」を設立し、設立趣旨に賛同する企業が参画している。提供データの扱いについては、①各企業が提供可能と判断したデータを利用するとともに、データを取り扱う行政側の職員を限定するほか、大学研究者らも国家公務員として任用し守秘義務の下で情報管理及び保秘の徹底、目的外利用の禁止、②参加企業間でのデータの共有は行わない、③成果物の公表に際しては参加企業の個別データの状況が明らかとならないようにし、参加企業の了解を得た上で行う、など協議会でのルールを定めている。(CTI)
- 神戸市においては、データホルダー企業と官側で連携協定を締結し、市政課題解決に向けたビッグデータを活用する取組(「都心・三宮の再整備の効果検討」等)が行われている。(神戸市)
- ウェブスクレイピングによる情報収集・統計作成に関して、データホルダーから承諾を受け、企業側から提示のあったアクセス時間帯・頻度の制限、IPアドレスの事前登録などの対応により、サーバーへの負荷を分散している。(CPI)

信頼関係の構築には、上記のように協定書を締結するなどオフィシャルな協力関係作りとともに、信頼関係に基づく民間側と行政側の窓口担当者間での密な連携が欠かせない。民間データを行政側で取扱うルールに関してはプロジェクト毎に協定書等を作成することが通例となっているが、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雛形を整備し、締結の実績を蓄積していくことは有効な取組である。

② データ提供者へのインセンティブ付与

民間データは該当企業のビジネスリソースとなっているものもあり、それを継続的に提供してもらうためのインセンティブに関して、考慮や検討が必要となる。

- 民間企業がビジネスとして使用しているデータについて、データの分析過程で出てきた副次的な情報をデータホルダーにフィードバックするなど、Win-Winの関係となれる利活用モデルを構築することで、価格面を含めた取得・利用要件を整えることが必要である。(メッシュ型流動人口検証WG)
- 自社のメリットというよりは、公的施策への貢献という企業側の社会貢献に対する意識に支えられて、取組は進められている。(CTI)

企業へのインセンティブとしては、データ提供の対価を支払うことが挙げられるが、公的機関が継続的に対応できるか難しい場合も考えられる。研究機関では対価として、データの分析結果や利活用性に関する研究レポートの発行など、データの品質を保証することにより、無償でのデータ提供を受けている事例もある。社会貢献事実の公表や政府による付加価値のフィードバックなどのイン

センティブも考慮した上で、適切な価格設定を官民が協議しながら、Win-Win の関係を構築する必要がある。

③ データ提供の中断に関するリスク対応

民間データは、企業の経営方針の変更や倒産などにより継続的にデータを受け取ることができなくなる可能性がある。またデータ購入にあたっては、購入価格の上昇や、競争入札によるデータの提供元の変更などにより統計の継続性を担保することが難しくなるなどのリスクがあるので、そういう事態が発生した時の対応を事前に検討しておく必要がある。

(3) ビッグデータを扱う人材・体制面

ビッグデータの活用を社会的に推進していくために、データを扱うスキルや知識を有する人材の育成を進めると共に、具体的にデータを活用し業務を遂行する組織では、そういった人材を確保し組織的にノウハウの蓄積を進めていく必要がある。

- 業務の内容・性質や、必要とされるハードウェア・ソフトウェア環境、及び人的リソースを考慮し、外部事業者との間で役割分担・委託を行う一方、スキル・ノウハウが特定の担当者に集中しすぎないようにするなど、組織的な蓄積が図れるよう取組を行っている。(C P I)
- 根拠に基づいた政策立案 (EBPM) を推進するため、データサイエンティストを非常勤嘱託職員として採用したほか、「神戸市データアカデミー」ではデータホルダーとも連携し、市職員のデータ活用リテラシー向上を目的とした人材育成に取り組んでいる。(神戸市)
- 職員が交替しても調査統計の品質が低下しないよう、「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⁸」を作成した。(パーソントリップ調査)

業務の品質を確保するために上記のような人材育成・確保やノウハウを継承する取組が行われているが、これらは政府と企業の信頼関係醸成・人脈作りにもつながっていくことが期待できる。人材交流については民間人材を行政において活用することだけでなく、行政職員が研究機関に出向し、民間のデータ活用ノウハウを持ち帰り、それを自府省に展開することも有効な手段であると考えられる。

(4) システム・技術面

ビッグデータを集計・加工するシステムは、データの種類やボリューム、取扱いの形態によって異なってくるが、それぞれ適時適切に扱うためのシステム等を整備する必要がある。

- 民間企業から提供されたデータに対して、セキュリティを確保しつつ効率的に大量のデータを扱うためのシステム環境を構築している。(C T I)
- ウェブ上にある不定型なデータを処理するために、データのフォーマット変換やエラーチェック、分類符号化の実施や、統計を作成するためのプログラ

⁸ 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>

ムを作成している。(C P I)

今回の事例はデータの処理に特化したものとなっているが、DXの時流に沿って全体最適のシステムとするためには、データ処理に閉じずに、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となる。また今回の取組を通じて、官民のデータ連携や、ウェブスクレイピング等への対応を見据えたウェブ掲載データ・フォーマットなど各種データの標準化等を進めていくことにより、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っていくことが期待される。

(5) その他（情報保護、法令関係など）

民間企業の視点からは、ビッグデータは商材そのものであったり、ビジネス競争力の源泉であったりするので、安心してデータを提供できるように、データの流通環境を整備する必要がある。また個人情報保護に関しては直接のデータ提供元である企業による担保をはじめ、個人情報をどう取り扱うべきかについて、社会的なコンセンサスの確立を図る必要がある。

また、ビッグデータ利活用に係る環境や技術の発展・変化にあわせ、公的制度の中で、ビッグデータの活用や扱いをどのように位置づけていくか、整備検討を進めることが重要である。

4 活用の推進に向けて

新型コロナウイルス対応において、モバイルデータを用いた外出自粛の状況把握が注目されるなど、ビッグデータ活用の動きが進展しつつあるが、一方で社会的にはプライバシーやガバナンス等、過剰なデータ流通に警鐘を鳴らす観点も増しているところである。社会経済のデジタル化が進む中で、ビッグデータの更なる活用が期待されており、活用における各種課題の解決に向け産官学で一層の連携を図る必要があると考えられる。

別紙

- 別紙 1 ビッグデータ連携会議の概要
- 別紙 2 ビッグデータ連携会議の開催実績
- 別紙 3 ビッグデータの公的統計への利活用事例等について

以上

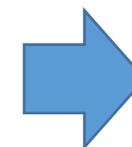
ビッグデータ連携会議の概要

主旨

- 「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を選定
 - ・関係者の協力を得て集中的に協議、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
 - ・各府省、地方公共団体、民間企業等における統計的分析や統計作成目的によるデータ等の相互利活用の推進

対象分野

統計分野	活用するビッグデータ例
物価	・POSデータ（商品別物価の把握） ・ポイントカードデータ ・ウェブサイトからの価格情報の自動収集 ・不動産価格
労働	・求人情報
消費	・POSデータ（商品別金額・購入数量） ・ポイントカード、クレジットカード ・家計簿ソフト
電力	・スマートメーター（全世帯・企業の30分ごとの電力使用量）
情報通信	・携帯電話通信データ
運輸	・道路の交通量データ ・自動車の位置情報



会議テーマ

- 府省による公的統計へのビッグデータ活用に関する検討事例の共有
- 民間ビッグデータの利活用や統計的分析事例の紹介
- 民間から提供を受けたデータを用いた分析の報告

（資料）経済同友会「豊かさの増進に向けた経済統計改革と企業行動～新たな指標群『GNIプラス』の提案～」（2016年9月）

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
<オブザーバー>
清家 武彦 日本経済団体連合会経済政策本部統括主幹

ビッグデータ連携会議の開催実績

資料5の参考
別紙2

開催日	テーマ名	概要
第1回 H30.5.23	ビッグデータの利活用に向けて	ビッグデータ連携会議の立ち上げ及びビッグデータの概要・先進的な事例に関する共有。
第2回 H30.6.27	労働市場の民間ビッグデータの経済統計としての活用可能性（R社）	R社が保有する 転職前後の給与 を使用した「 転職時の賃金変動状況 」の紹介。
第3回 H30.7.25	携帯GPSを活用した流動人口把握による統計サービスの向上（A社）	A社が保有する 携帯アプリのGPS を使用した 流動人口データ （メッシュ型 & ポイント型）の特徴や活用事例の紹介。 メッシュ型流動人口検証WG の発足。
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ビッグデータ活用について（Z社）	人口、所得、店舗など必要なデータを 地図上に取り込んで可視化 し、ビジネスに活用していく エリアマーケティング におけるZ社によるコンサルティング事例の紹介。
第5回 H30.12.5	PT調査（国土交通省）、CTI（総務省統計局）におけるビッグデータの利活用について	携帯基地局情報・交通系ICカード情報 等を活用した パーソントリップ調査 や、 POSデータ・クレジットカード 等を使用した 消費動向指数 に関する取組状況の紹介。
第6回 R1.5.22	メッシュ型流動人口検証WG報告、ビッグデータを活用した新指標開発事業について（経済産業省）	メッシュ型流動人口検証WG の結果報告、及び、SNSの書込ログを活用した新指標開発、 POSデータ を活用した 商業動態統計調査（試験調査：家電大型専門店分野） に関する取組状況の紹介。
第7回 R1.7.31	電力データ活用の取組（G事業組合、A区）、労働BDの統計的利活用検討（R社）	G事業組合による スマートメーター の情報を活用した、民間の取組や、A区での 自治体業務への活用 可能性に関する検討事例の紹介。 R社による 民間の転職情報 を活用した統計作成の取組の紹介。
第8回 R1.11.13	CPIへのウェブスクレイピングの活用について（総務省統計局）	外国パック旅行費 等に関して WEBスクレイピング により価格情報を収集し、消費者物価指数を作成する取組の紹介。
第9回 R2.2.26	自治体におけるビッグデータ活用の取組（K市）、民間ソリューションの紹介（Y社）	K市 基幹駅周辺の都市開発 における流動人口データを始めとしたビッグデータの活用事例と、K市でのその他の取組を紹介。 Y社の保有する デスクリサーチツール を中心としたサービスの紹介。
第10回 R2.9.30	地球観測データを活用したSDG指標の算出（JAXA）	JAXAが所有する 衛星観測データ を用いた SDG15.4.2（山地植生被覆指数） の算出に関する取組の紹介。 観測データ利活用検証WG の発足。

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マイクロ)、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、第三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。

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	<p>主目的</p> <p>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。</p>	<p>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。 (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。)</p>	<p>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。</p>	<p>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。</p>
<p>副次効果</p> <p>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</p>	<p>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。)</p>	<p>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。</p>	<p>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準誤差の縮減につながる。</p>	<p>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。</p>
	<p>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。</p>			<p>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。</p>
課題・取組	<p>ビッグデータの選定・加工</p> <p>【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・外国バック旅行を対象とした。</p> <p>【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 ・外国バ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</p>	<p>【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。</p>	<p>【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。</p>	<p>【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。調達は安価で協力してくれることを重視して選定した。</p> <p>【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、経由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)</p>

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】 ビッグ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>

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) POSデータ	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。
その他				

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
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/kajgi/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