

ビッグデータの更なる活用の方向性 ～政策の質の向上を目指して～

令和4年（2022年）6月2日
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決定

1. 取りまとめの経緯

「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」（以下「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（2018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、平成30年（2018年）5月23日より令和4年（2022年）6月2日までの間に毎年4回程度、計18回開催してきた。

その中で、令和2年（2020年）12月23日のビッグデータ連携会議では、ビッグデータの活用法やメリット、既存の公的統計の中でビッグデータを活用する場合を念頭に置いた活用上の課題とその解決事例及び今後対応していく方向性などの整理を行った²。その後、令和3年（2021年）8月27日の第17回統計委員会企画部会における統計法施行状況に関する審議において、このような令和2年（2020年）12月に整理した内容を報告したところ、「ビッグデータの既存の公的統計への活用と、既存の公的統計では把握することのできなかつた新たな指標の開発については、性質が異なるため分けて考えるべきではないか。」との指摘をいただいた。

また、令和3年（2021年）6月23日のEBPM推進委員会データ利活用ワーキンググループの「取りまとめ」においては、EBPMが実際の政策の質の向上に結び付いていくためには、行政が、民間データを含めた様々なデータを整備又は取得し、それぞれのデータの特徴を踏まえた利活用を図ることが必要である旨の指摘がされた。

本資料は、令和2年（2020年）12月の整理以降のこうした指摘事項も踏まえ、ビッグデータの活用を更に推進し、政策の質の向上等を図るため、ビッグデータの活用事例について、既存の公的統計との関係性の観点から再整理を行い、今後のビッグデータの更なる活用に向けた方向性を示すものである。

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

² 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：https://www.soumu.go.jp/main_content/000732175.pdf

2. ビッグデータの活用事例の再整理

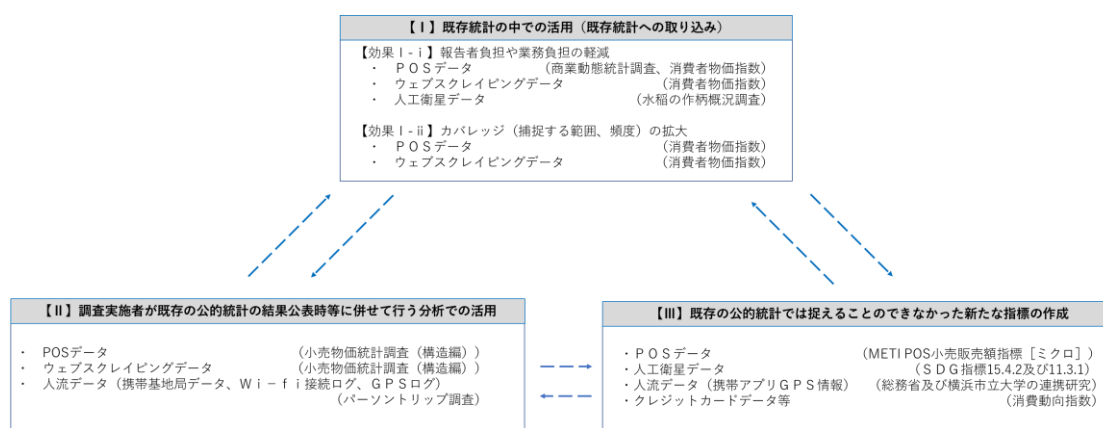
令和2年(2020年)12月の整理においては、ビッグデータの活用法やメリットについて、「(1)既存統計の補完」と「(2)報告者負担や業務負担の軽減」の2つに分け、「(1)既存統計の補完」の中に「④新たな指標等の作成」を含めていた。その後、この点について、複数の有識者から、「ビッグデータの既存の公的統計への活用と新たな指標の開発は分けて考えるべきではないか」との指摘を受けたところである。

このような指摘を踏まえ、本資料では、ビッグデータの活用事例を、既存の公的統計との関係性及び活用による効果の観点から、

【Ⅰ】 既存の公的統計の中での活用（既存統計への取り込み） 活用による効果Ⅰ-i 報告者負担や業務負担の軽減 活用による効果Ⅰ-ii カバレッジ（捕捉する範囲、頻度）の拡大
【Ⅱ】 調査実施者が既存の公的統計の結果公表時等に併せて行う分析での活用
【Ⅲ】 既存の公的統計では捉えることのできなかつた新たな指標の作成

として再整理することとした。なお、これらは便宜的に分類を行ったものであり、必ずしも普遍的な分類ではなく、中・長期的には、活用状況の変化や活用技術の発展等により、これらの個別の活用事例が別の区分に分類される可能性があることにも留意が必要である。

ビッグデータの活用事例の整理（イメージ）



【I】既存の公的統計の中での活用（既存統計への取り込み）

ビッグデータを既存の公的統計の中に取り込んで活用するためには、活用するビッグデータの特性や既存の公的統計との関係性等を十分に把握すること、安定的・継続的なデータの入手経路を確立することなどが必要不可欠である。現時点においては、「POSデータ（販売時点情報）」、「ウェブスクレイピング技術を活用して大量に取得したウェブ情報」（以下「ウェブスクレイピングデータ」という。）及び「人工衛星データ」の3種類のデータが既存の公的統計の中で活用されている。

効果 I-i 報告者負担や業務負担の軽減

新たなデータソースとしてビッグデータを活用し、既存の統計データや統計調査業務を代替することにより、一部の事項について調査する必要がなくなる、従来の調査員等による人手を要する調査が不要となるなど、報告者負担の軽減や業務負担の軽減が可能となる。

- 消費者物価指数（CPI）では、ウェブスクレイピング技術³を活用して自動的に大量のウェブ上の価格情報を収集することにより、該当する品目に係る価格の報告を不要にするとともに、行政による効率的なデータ収集を実現している。また、POSデータを活用し、該当する品目に係る価格調査を廃止することにより、報告者負担及び業務負担の軽減を行っている。
(総務省)
- 商業動態統計調査では、家電大型専門店に関する調査についてPOSデータによる提出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。
(経済産業省)
- 水稲の作柄概況調査では、人工衛星データを用いて現地での水稲の実測調査の代替を行うことにより、実査の負担を軽減している。 (農林水産省)

効果 I-ii カバレッジ（捕捉する範囲、頻度）の拡大

高頻度かつ多種大量に生成されるビッグデータの特性を活かすことにより、既存統計のカバレッジ（捕捉する範囲や頻度）を拡大することが可能となる。

- 消費者物価指数（CPI）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」及び「国内宿泊料」について、ウェブスクレイピングデータの活用により、旅行会社・プラン数や路線数の把握対象の拡大、高頻度（毎日）の価格把握などを行っている。また、POSデータの活用により大量の価格データを取得・活用することで、統計の精度向上に努めている。
(総務省)

³ ウェブスクレイピング技術：ウェブサイトから情報を抽出するコンピュータソフトウェア技術

【Ⅱ】調査実施者が既存の公的統計の結果公表時等に併せて行う分析での活用

ビッグデータの活用については、「【Ⅰ】既存統計の中での活用（既存統計への取り込み）」のように既存の公的統計の中で活用することのほか、調査実施者が既存の公的統計の結果公表時等に併せて行う分析において活用する取組も行われている。統計調査の結果公表に際し、ビッグデータを活用した分析結果を併せて示すことにより、より多角的な分析が可能となる。

- パーソントリップ調査では、必要な情報について統計的に結果精度を確保できるように調査を実施した一方で、人流データ（携帯基地局データ等）を活用してパーソントリップ調査では把握できない人の動きを把握している。
（国土交通省）
- 小売物価統計調査（構造編）では、既存の調査の一部を中止し、POSデータを活用した分析に移行した。併せて、ウェブスクレイピング技術を活用し、ネット価格の分析を実施することとしている⁴。
（総務省）

⁴ 店舗形態別価格調査及び銘柄別価格調査について、2021年12月の調査を最後に中止し、POSデータを用いた店舗形態別価格等に関する分析に移行している。

【Ⅲ】既存の公的統計では捉えることのできなかつた新たな指標の作成

既存の公的統計に比べはるかに早い時点で入手可能となる、従来の手法では取得できなかった大量のデータが利用可能となるなどのビッグデータの特徴を活かし、既存の公的統計では捉えることのできなかつた新たな指標等を作成することについて試行や研究が進められている⁵。新たな指標としては様々なものが考えられるが、例えば、既存の公的統計の先行指標を作成する場合には、既存の公的統計の結果との相関を確認するなど、有用性の検証が必要となることに留意する必要がある。

- 宇宙航空研究開発機構（JAXA）では、人工衛星データを用いて、国連持続可能開発目標（SDGs）指標の開発・検証を行っている。SDG 指標 15.4.2（山地グリーンカバー指数）に関する検証結果について、令和3年6月のSDGs推進本部幹事会において報告を行うとともに、連携会議レポートとしてHPに公開している⁶。また、指標 11.3.1（人口増加率と土地利用率の比率）についても検証を進めている。（宇宙航空研究開発機構（JAXA））
- 経済産業省では、HP（BigData-STATS ダッシュボード（β 版）⁷）を作成し、POSデータを活用して試験的に新たな指標（METI POS 小売販売額指標 [ミクロ]）を公開している。（経済産業省）
- 消費動向指数（CTI）では、クレジットカードデータ等の民間ビッグデータを活用し、従来よりも速報性に優れた指標を検討・開発中である。（総務省）
- 総務省及び横浜市立大学では、人流データ（携帯アプリ GPS 情報）と宿泊旅行統計調査の延べ宿泊者数との比較検証を行い、速報性の高い指標の開発について研究を進めている⁸。（総務省及び横浜市立大学の連携研究）

⁵ 月例経済報告等に関する関係閣僚会議においても、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（JCB消費NOW：<https://www.jcbconsumptionnow.com/>）等が活用されている。

⁶ ビッグデータ連携会議レポート：「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」の議論を踏まえ、執筆者による調査・研究の成果をまとめたもの。

No.1 SDG15.4.2（山地グリーンカバー指数）の検証

日本語版 https://www.soumu.go.jp/main_content/000763967.pdf

英語版 https://www.soumu.go.jp/main_content/000763968.pdf

⁷BigData-STATS ダッシュボード（β 版）：https://www.meti.go.jp/statistics/bigdata-statistics/bigdata_pj_2019/index.html ※令和4年（2022年）4月1日公表をもって更新を終了している。

⁸令和2年（2020年）7月より、総務省と横浜市立大学が連携して研究を進めている。

3. ビッグデータの種類別の整理

「2. ビッグデータの活用事例の再整理」をビッグデータの種類別に行ってみると、以下の表のとおりとなる。

「POSデータ」、「ウェブスクレイピングデータ」及び「人工衛星データ」については、「【II】調査実施者が既存の公的統計の結果公表時等に併せて行う分析での活用」及び「【III】既存の公的統計では捉えることのできなかった新たな指標の作成」に分類される事例があるほか、「【I】既存の公的統計の中での活用（既存統計への取り込み）」に分類される事例も見られる。これらのデータについては、活用するビッグデータの特性・既存の公的統計との関係性の把握や、安定的・継続的なデータの入手経路の確立が比較的進んでいることから、既存の公的統計の中での活用が行われていると考えられる。

一方で、「人流データ」及び「クレジットカードデータ等」については、「【II】調査実施者が既存の公的統計の結果公表時等に併せて行う分析での活用」及び「【III】既存の公的統計では捉えることのできなかった新たな指標の作成」に分類される事例はあるものの、現時点では「【I】既存の公的統計の中での活用（既存統計への取り込み）」に分類される事例は見られない。

< 既存の公的統計との関係性（ビッグデータの種類別の整理） >

ビッグデータの種類	活用方法とメリット			
	【I】既存の公的統計の中での活用 (既存統計への取り込み)		【II】調査実施者が既存の公的統計の結果公表時等に併せて行う分析での活用	【III】既存の公的統計では捉えることのできなかった新たな指標の作成
	I - i 報告者負担や業務負担の軽減	I - ii カバレッジ（捕捉する範囲、頻度）の拡大		
POSデータ	消費者物価指数(CPI)、商業動態統計調査	消費者物価指数(CPI)	小売物価統計調査（構造編）	METI POS小売販売額指標 [ミクロ]
ウェブスクレイピングデータ	消費者物価指数(CPI)	消費者物価指数(CPI)	小売物価統計調査（構造編）	—
人工衛星データ	水稻の作柄概況調査	—	—	SDG指標15.4.2及び11.3.1
人流データ	—	—	パーソントリップ調査（携帯基地局データ等）	総務省及び横浜市立大学の連携研究（携帯アプリGPS情報）
クレジットカードデータ等	—	—	—	消費動向指数(CTI)

4. ビッグデータの活用における課題と考察

ビッグデータについて、「【Ⅰ】既存の公的統計の中での活用（既存統計への取り込み）」を行う場合には、結果精度を確保し、安定的・継続的に結果を提供し続ける必要があることから、以下のような点について検討・整備を進めることが必要となる⁹。

- ① ビッグデータの特性把握
ビッグデータはデータ生成方法や生成プロセス等に起因する偏りや特徴を有しているため、既存の公的統計の中での活用にあたっては、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検証するなど、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。
- ② ビッグデータの安定的・継続的な入手方法の確立
協定書の締結に基づくオフィシャルなルール設定を含めたデータホルダーとの間のデータ取扱いのルールの設定、データ提供の対価の支払いを含めたインセンティブの付与、データの提供が中断することに対するリスク対応等について検討すること。
- ③ ビッグデータを扱う人材・体制面の整備
具体的にビッグデータを活用し、業務を遂行する組織では、ビッグデータを扱うスキルや知識、データホルダーとの調整能力等を有する人材を確保し、組織的にノウハウを蓄積、継承していくこと。
- ④ システム・技術面の整備
取り扱うビッグデータの種類やボリューム、取扱いの形態にも応じて、それぞれ適切に扱うためのシステム等を整備すること。

一方で、「【Ⅱ】調査実施者が既存の公的統計の結果公表時等に併せて行う分析での活用」を行う場合には、安定的・継続的な入手方法の確立や人材・体制、システム面の整備などを事前に進めておく必要は必ずしもなく、アドホックに分析を行い、既存の公的統計の調査結果を公表する際に、ビッグデータによる分析結果も参考として併せて提供することも有効と考えられる。

また、「【Ⅲ】既存の公的統計では捉えることのできなかつた新たな指標の作成」を行う場合には、新たな指標の性質にもよるが、ビッグデータの特性等が十分に把握し切れていなかったとしても、ビッグデータには偏りがあり得ることに留意した上で指標を開発し、試験的な指標を公開することも有効と考えられる。その際、データの特性や加工方法など、公開可能な情報についてはできる限り公開することが望ましい。

⁹ 詳細については、「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-」（令和2年12月23日ビッグデータ連携会議決定）の「3 ビッグデータの活用における課題及びその解決方法」を参照。

ビッグデータについて、「【Ⅰ】既存統計の中での活用（既存統計への取り込み）」を行うためには、上記のように様々な事項についてあらかじめ検討する必要があり、新たなビッグデータを既存の公的統計の中で活用するには課題解決のために多くの困難が伴うと考えられる。日本において、公的統計におけるビッグデータの活用の取組が全体的になかなか進んでいかない印象があるのは、これまで、このように厳しい条件での活用を目指して取組が進められてきたことにも原因があるのではないかと考えられる。一方で、データの利活用により政策の質の向上等を図るためには、公的統計に限らず、民間データを含めた様々なデータについて、それぞれのデータの特徴を踏まえつつ利活用を進めていくことが重要である。

以上を踏まえると、今後、ビッグデータの更なる活用を推進していくためには、「【Ⅰ】既存の公的統計の中での活用（既存統計への取り込み）」のみでなく、より多角的な分析を可能とする「【Ⅱ】調査実施者が既存の公的統計の結果公表時等に併せて行う分析での活用」及び速報性などに優れた「【Ⅲ】既存の公的統計では捉えることのできなかつた新たな指標の作成」の取組も併せて推進していくことが有効な手段であると考えられる。これにより、ビッグデータの活用の裾野が徐々に広がり、ノウハウや事例の蓄積、各種データの検証等が進んでいくものと考えられる。

5. ビッグデータの更なる活用に向けた方策

「4. ビッグデータの活用における課題と考察」を踏まえると、今後、ビッグデータの活用を更に推進していくための具体的な方策として、以下が考えられる。

(1) ビッグデータ・トライアル（ビッグデータを活用した試行的な取組）の実施及び成果の公開、それを推進するための環境整備

ビッグデータを公的統計の中で活用し、定期的に公表していくためには、ビッグデータの特性を把握し、安定的・継続的なデータの入手方法を確認し、人材面・体制面、システム面の整備を行うなど、様々な取組が必要となる。そのための第一歩として、ビッグデータ・トライアル（ビッグデータを活用した試行的な取組）を順次実施し、その成果（β版、試算値、参考値など）を公開し、フィードバックを得て更に検証を進めていくことが、公的統計におけるビッグデータの活用に向けた有効な取組であると考えられる。こうした試行的な取組を行う環境を整備し、これらの取組を推進するため、総務省が中心となって、新たに生成され、活用が可能となったデータを含めた様々なビッグデータを試験的に利用すること（ビッグデータ・シェアリング）ができる場や、ビッグデータを活用した取組の成果を公開する場（ビッグデータ・ポータル）を準備することも有益と考えられる。

なお、公開に際しては、結果の再現性やプロセスの透明性を確保し、公的統計の活用につなげていく観点からも、活用したビッグデータの出所やその生成方法、データの特性やデータの処理方法等について可能な限り情報を公開することが望ましい。また、公開方法については、ビッグデータは基本的に統計の作成を目的として生成されたものではないことを踏まえ、データの種類やその成果の質等に応じ、公的部門以外の研究機関・民間企業などから公開する、共同研究を行った複数機関の連名で公開するなど、様々な方法が考えられる。これらの取組を進めるに当たっては、後記「(4) 関係機関、産官学との連携」のとおり、関係者間の連携が必要不可欠である。

(2) 既存の公的統計の中で活用されているビッグデータの更なる活用の推進

既存の公的統計の中で活用されている「POSデータ」、「ウェブスクレイピングデータ」及び「人工衛星データ」については、調査項目の代替を含めた報告者負担や業務負担の軽減等の効果が期待される。また、統計の作成プロセスを透明化し、人の手を介さずに行うことによってヒューマンエラーの発生を防止し、データの信頼性を向上させる効果も期待される。これらのデータについては、既存の公的統計における必要性や費用対効果等も踏まえ、活用の対象を拡大するなど、更なる活用が望まれる。一方、これら以外のデータの公的統計における活用に向けては、まずは、実証研究や上記のビッグデータ・トライアルにおいて、そのデータの特性や質を確認・評価していくことが必要である。

(3) 外部の視点等の積極的な取り込み

ビッグデータ等の活用は新たな取組であり、既存の枠組みにとらわれない外部の視点等を積極的に取り入れていくことが重要である。また、新たな指標の検討等に当たっては、データの分析方法等を機動的に改善・修正していくアジャイル型の視点や、ビッグデータの活用により何が可能になるのかといった結果を見据えた視点、社会経済情勢が変化する中で、どのようなデータが求められているかといったニーズ把握の視点も重要である。ただし、これらの視点を踏まえて指標の作成や分析方法等を変更した場合にも、利用者に配慮し、可能な限りそれらに関する情報を公開することが望ましい。

(4) 関係機関、産官学との連携

社会経済のデジタル化が進展する中、ビッグデータの更なる活用が期待されており、公的機関等をはじめとする様々な機関においてデジタル化の推進やビッグデータの活用に向けた取組が行われている。一方で、データホルダーからは、自身の保有するビッグデータをどのように活用したら良いか分からない、統計技術的な知識が不足している、といった声も寄せられている。「(1)ビッグデータ・トライアル（ビッグデータを活用した試行的な取組）の実施及び成果の公開、それを推進するための環境整備」をはじめ、ビッグデータの更なる活用の推進に当たっては、公的機関、データホルダー、研究機関・有識者など、利活用に係る関係者を結びつけて情報共有を行い、互いに連携することが重要である。また、各種課題の解決に向けて、総務省が中心となって各府省や関係機関とも連携しつつ、各種データの実証研究等を行い、ビッグデータ連携会議に報告して情報の共有・データの評価を行っていくなど、引き続き産官学で一層の連携を図っていく必要がある。

以上

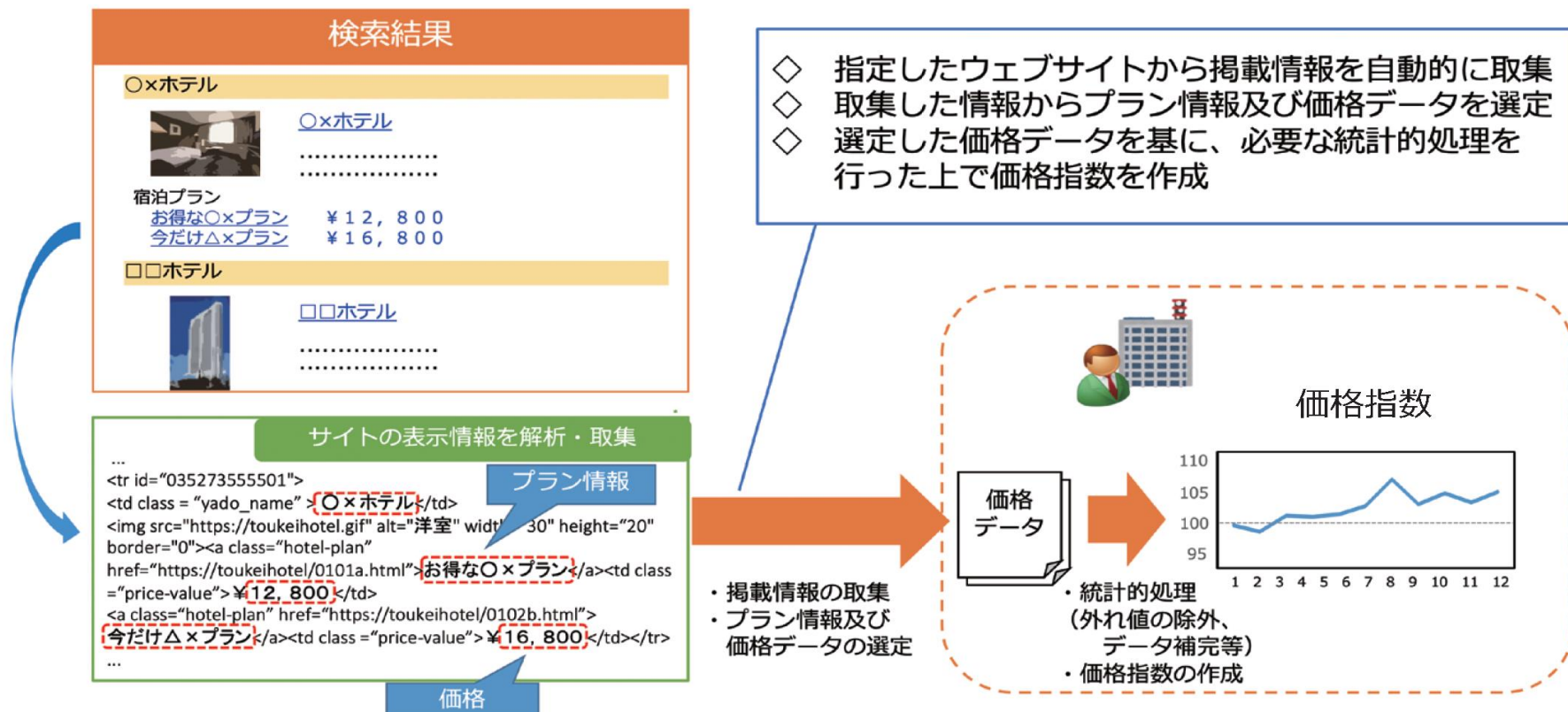
別紙 ビッグデータの公的統計への利活用事例等について

参考 1 ビッグデータ連携会議の概要

参考 2 ビッグデータ連携会議の開催実績

- 総務省の消費者物価指数(CPI)において、
 - ① POSデータ（販売時点情報）の活用、
 - 〔 7品目（パソコン（デスクトップ型）、パソコン（ノート型）、タブレット端末、カメラ、テレビ、ビデオレコーダー、プリンタ） 〕
 - ② ウェブスクレイピング技術を活用したHPからの大量の自動的な価格収集（3品目（航空運賃、宿泊料、外国パック旅行費））を実施している。

「ウェブスクレイピング」技術の活用イメージ（宿泊料）



消費動向指数（CTI）の概要

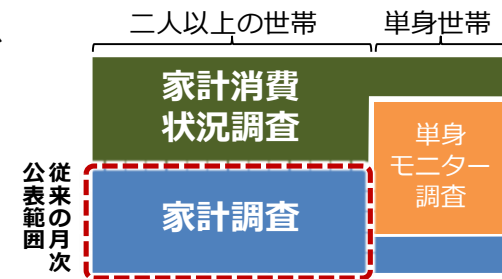
- ビッグデータ等を活用し、消費動向をマクロ・ミクロの両面から捉える速報性の高い消費指標の体系：消費動向指数（CTI：Consumption Trend Index）を新たに開発
- 平成30年1月分から参考指標として公表開始

※データソースは、当初は既存統計を利用。研究分析・検証を経た後、ビッグデータを順次活用

世帯消費動向指数 （CTIミクロ）

世帯の平均消費支出額（10大費目別、世帯類型別など）の月次動向を示す統計指標

- ◆ 家計調査（標本規模：二人以上の世帯 約8千、単身世帯 約7百）の結果を、
 - 家計消費単身モニター調査（標本規模：2千4百）
 - 家計消費状況調査（標本規模：約3万）
 の結果等と統計的手法によって補正・補強し、標本規模を擬似的に拡大、推計精度を向上

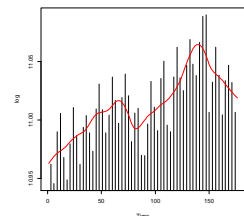


総消費動向指数 （CTIマクロ）

国内経済における個人消費総額（GDPにおける家計最終消費支出）の月次動向を示す統計指標

- ◆ GDP統計（家計最終消費支出）をターゲットとして、最新の動向を推測
- ◆ GDP統計の四半期別公表値では観測できない月次の値を時系列分析の手法と各種のデータを用いて推計※

※世帯消費動向指数（総世帯）、サービス産業動向調査結果（サービス産業計）、商業動態統計調査結果（小売業計）、第三次産業活動指数（広義対個人サービス）、鉱工業生産指数（消費財計）を説明変数として使用



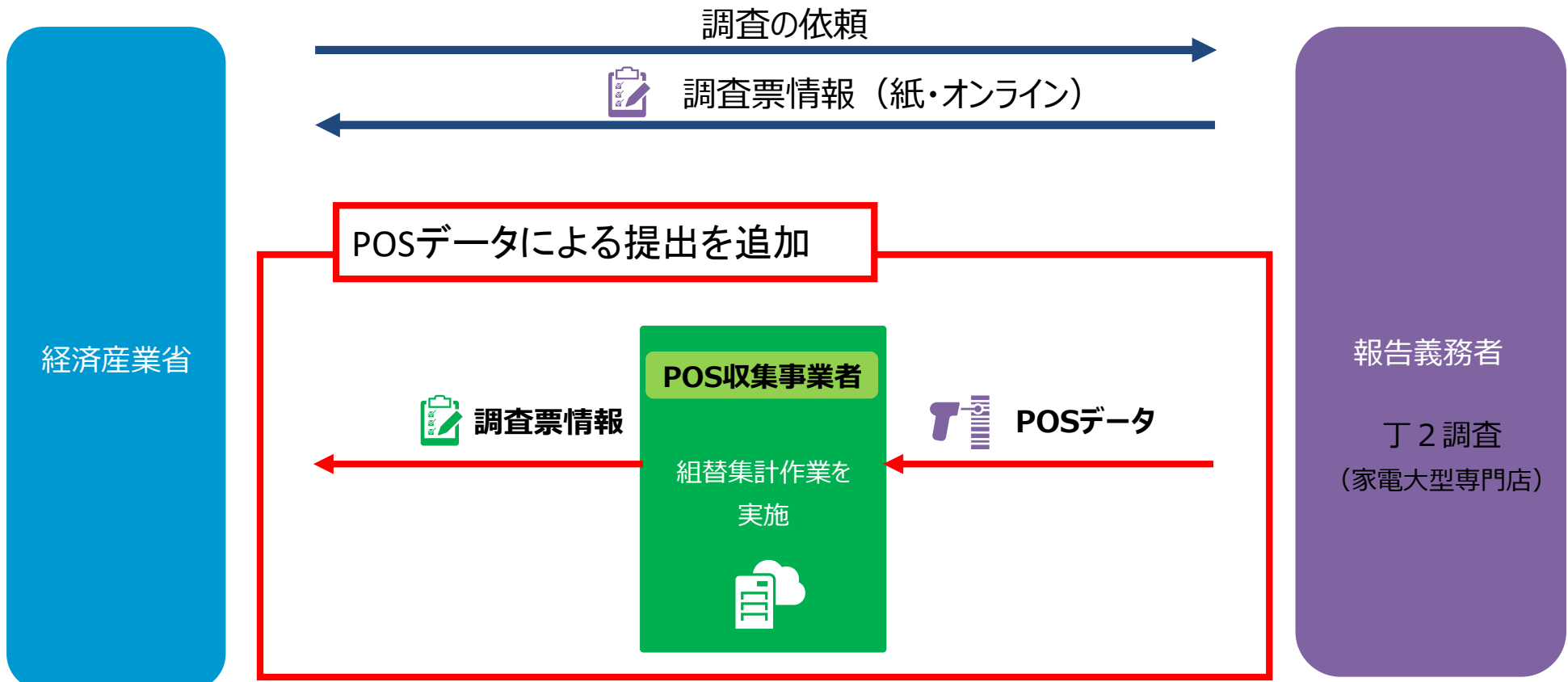
商業動態統計におけるPOSデータ活用

＜商業動態統計調査の概要＞

全国の商業（卸売業・小売業）の動向を把握するため、商業を営む事業所や企業の販売額等を毎月調査。

- ・調査対象：卸売業・小売業（百貨店、スーパー、コンビニエンスストア、家電大型専門店等）に属する事業所及び企業（約2万5千）。
- ・調査事項：商品販売額、従業者数等。
- ・調査結果は、月例経済報告、四半期別GDP速報、景気動向指数等に幅広く利用されている。

- 上記の統計調査のうち、「家電大型専門店」に関して、紙やオンラインに加えて、POSデータ（販売時点情報）を用いた提出方法を追加。



METI POS小売販売額指標 [ミクロ] (BigData-STATSダッシュボード(β版))

➤ データ出典

家電大型専門店：

ジーエフケーマーケティングサービスジャパン株式会社

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：

インテージ株式会社

➤ データ期間

販売金額指数：

週次（月曜～日曜） 2012年1月2日の週～最新週まで

月次（1日～月末） 2012年1月～最新月まで

前年比（%）：

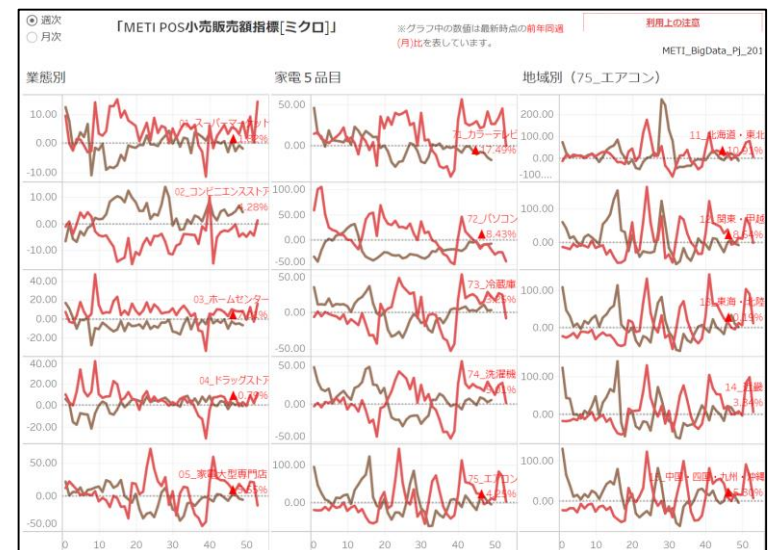
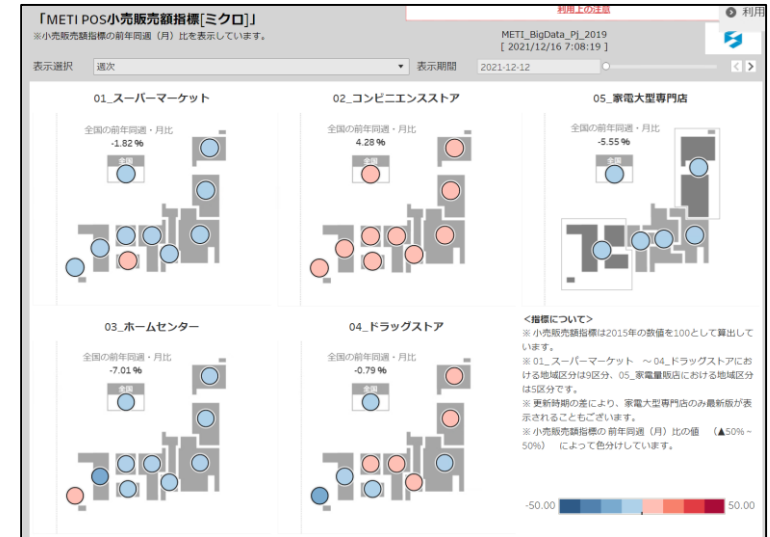
週次（月曜～日曜） 2012年12月31日の週～最新週まで

月次（1日～月末） 2013年1月～最新月まで

➤ 公表日

週次（月曜～日曜）：翌週金曜日 12時頃

月次（1日～月末）：翌月第1もしくは第2金曜日 12時頃



第6回東京都市圏パーソントリップ調査報告書におけるビッグデータの活用事例

- パーソントリップ調査では把握していない、「休日」(左図)及び「外国人」(右図)の活動状況を携帯電話基地局データ、wi-fi接続ログ、GPSログから集計・活用している。

ビッグデータからみた都市圏外居住者・休日の活動

- ・ 携帯基地局によるビッグデータより、平日と休日の発生集中量を比較すると、休日の方が平日よりも鉄道沿線外まで広く広がっている。
- ・ 都市圏外居住者は、都市圏内居住者と同様に都心部から鉄道沿線地域に集まっており、平日よりも休日の方が多く集まっている。

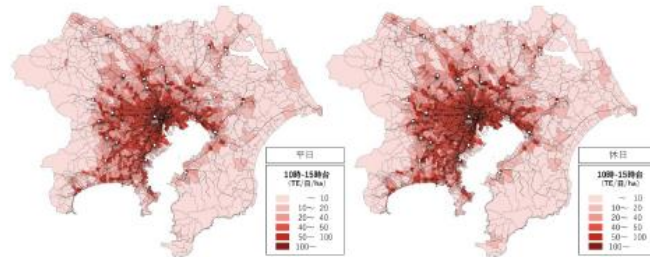


図 2-28 都市圏内居住者の集中交通量

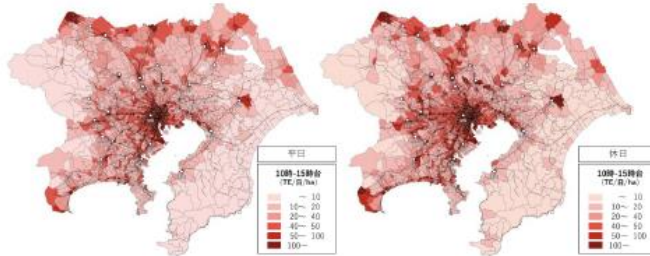


図 2-29 都市圏外居住者の集中交通量

資料：softbank ビッグデータより

<ビッグデータの概要>

データの種類	Softbank 携帯電話基地局データ
集計データ	東京都市圏内の小ゾーン間 OD ※都市圏内外・外内のトリップも集計対象
集計対象サンプル	東京都市圏内外居住者両方含む
データ期間	2018年10月20日(土)～26日(金)

ビッグデータからみた外国人の活動

- ・ アプリのログによるビッグデータより、外国人の滞在場所を見ると、東京駅から新橋駅の周辺、上野駅から秋葉原駅の周辺、浅草・スカイツリー周辺、池袋駅、新宿駅、渋谷駅、お台場などに集まっている。

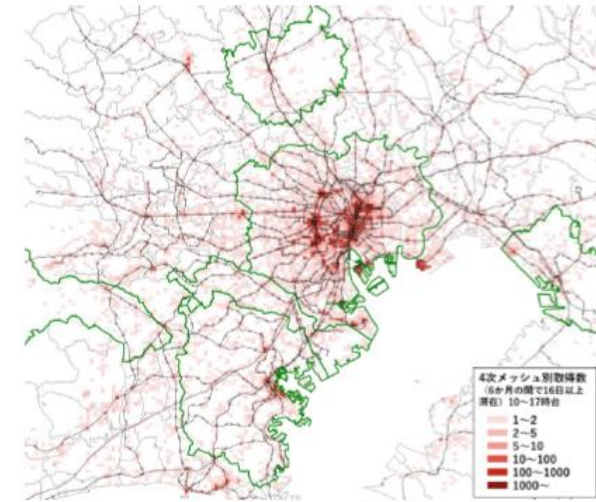


図 2-30 外国人の滞在場所

資料：ワイヤ・アンド・ワイヤレス (Wi2) Travel Japan Wi-Fi データより

<ビッグデータの概要>

データの種類	ワイヤ・アンド・ワイヤレス (Wi2) 「Travel Japan Wi-Fi」というアプリを利用した外国人の wi-fi 接続ログ、GPS ログのデータ
集計データ	東京都市圏内の4次メッシュ別の滞在場所
集計対象サンプル	滞在ログが15日以内の外国人
データ期間	2018年7月1日～2018年12月31日までの6か月
移動滞在の判定	4次メッシュ間の移動速度が20km/h未満を滞在と判定

小売物価統計調査（構造編）におけるビッグデータ活用

【小売物価統計調査の概要】

国民の消費生活上重要な支出の対象となる商品の小売価格及びサービスの料金を調査し、消費者物価指数や、その他物価に関する基礎資料を得ることを目的とする。調査対象は、商品の販売又はサービスの提供を行っている店舗・事業所。

【調査の見直し】

物価の毎月の動向を明らかにする「動向編」のほか、「構造編」として、地域別価格差調査、店舗形態別価格調査及び銘柄別価格調査を実施していたが、店舗形態別価格調査及び銘柄別価格調査については、2021年12月を最後に中止し、POSデータを活用した分析集計に移行。併せて、ネット価格の分析を実施。

◆2021年12月以前

区分		分析対象
調査員調査	地域別 価格差調査	都道府県別等
	店舗形態別 価格調査	店舗形態別
	銘柄別 価格調査	銘柄別

◆見直し後

区分		分析対象	データの概要
調査員 調査	地域別 価格差調査	(同左)	—
POSデータ		店舗形態別	食品・日用品等を取り扱う全国の 小売店舗から収集した、店舗形態 別販売実績 (店舗形態間で比較可能な品目・ 商品 (15品目・150商品程度) を毎年選定)
		銘柄別	家電製品を取り扱う全国の小売店 舗から収集した販売実績 (3～4品目を毎年選定)
ネット価格		店頭販売と ネット販売の 比較	ネット販売サイト等から得られる情報 (ウェブスクレイピング) (対象品目は毎年選定)

- 令和2年7月より、総務省と横浜市立大学が連携して研究を実施。
- 携帯アプリGPSの位置情報から作成した「メッシュ型人流データ」の実証研究を行い、「宿泊旅行統計調査」の日本人延べ宿泊者数の先行指標の開発や、指標を市区町村単位化する詳細化などを研究している。

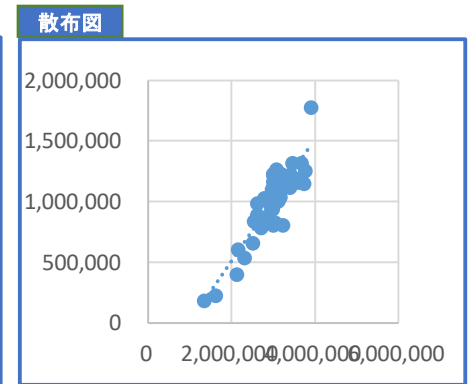
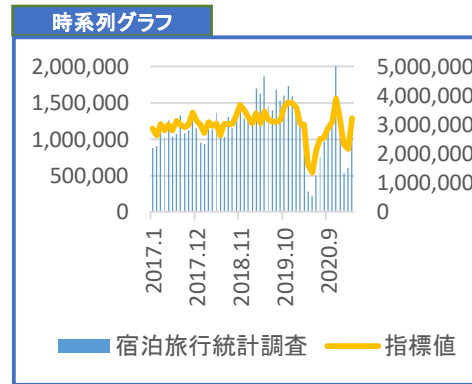
メッシュ型流動人口データ

- ①メッシュコード
- ②年月
- ③平日・休日別
- ④時間帯
- ⑤推計人口

CSVファイル

メッシュID	年	月	休日/平日	時間帯	人口	都道府県コード	跨り県数
623927494	2021	4	0	10	4	1	1
623927591	2021	4	0	0	140	1	1
623927591	2021	4	0	1	188	1	1
623927591	2021	4	0	2	127	1	1
623927591	2021	4	0	3	206	1	1
623927591	2021	4	0	4	297	1	1
623927591	2021	4	0	5	255	1	1
623927591	2021	4	0	6	234	1	1
623927591	2021	4	0	7	231	1	1

時系列分析
相関分析
回帰分析



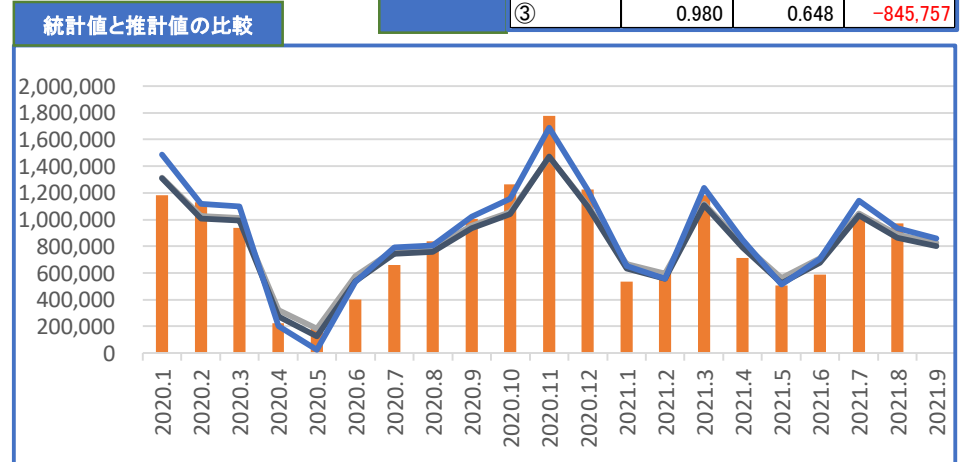
推計

相関係数
回帰式

	相関係数	傾き	切片
①	0.876	0.501	-489,228
②	0.938	0.524	-574,927
③	0.980	0.648	-845,757

宿泊旅行統計調査 (第2表、参考第6表・第8表)

施設所在地(47区分 及び運輸局等)	延べ 宿泊者数 1)	従業者数(4区分)、宿泊目的割合(2区分)				
		観光目的の 宿泊者が 50%以上		0~9人		
		観光目的の 宿泊者が 50%未満	1)	観光目的の 宿泊者が 50%以上	観光目的の 宿泊者が 50%未満	
令和3年8月	30,975,520	16,154,980	14,775,440	5,931,760	3,531,710	2,383,640
01北海道	2,203,820	1,243,240	959,340	449,800	274,820	173,740
02青森県	353,020	114,180	238,840	81,780	23,060	58,720
03岩手県	392,870	163,240	229,620	67,530	25,070	42,460
04宮城県	619,880	229,350	390,530	98,940	56,260	42,680
05秋田県	248,670	110,290	138,380	33,450	13,690	19,760
06山形県	401,210	196,380	204,830	122,270	78,830	43,440
07福島県	715,290	276,130	439,170	240,450	64,330	176,120
08茨城県	353,940	86,700	267,240	66,600	18,600	48,000
09栃木県	686,410	469,880	216,530	124,160	99,870	24,290
10群馬県	522,270	363,790	158,490	85,730	72,510	13,220
11埼玉県	331,290	88,050	243,240	36,900	28,980	7,910
12千葉県	1,423,640	911,970	508,030	237,090	131,060	106,040
13東京都	3,361,810	1,348,510	1,990,400	385,430	195,780	177,310



人工衛星からの取得データを利用した水稲作柄予測手法の導入

- 水稲の作柄予測について、気象データに加え、人工衛星データを利用することにより、調査精度を維持しつつ、実測調査に係る業務を合理化する手法の開発を平成29年度から2カ年かけて行い、令和2年度から運用を開始。

これまでの実測調査

職員等による実測作業の例

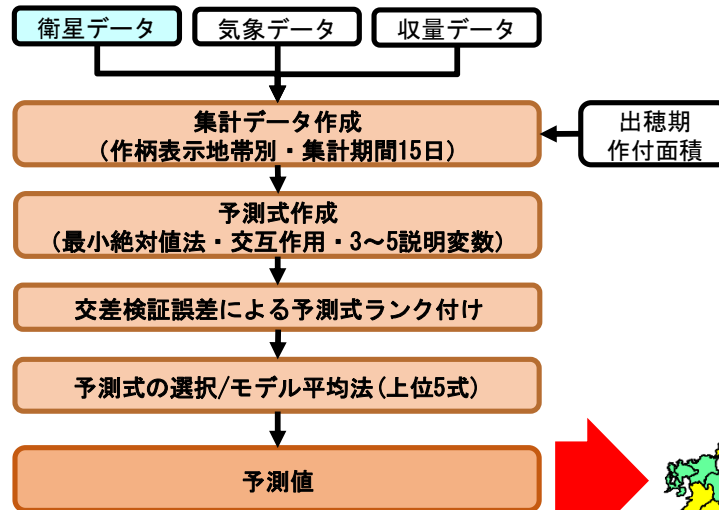


実測調査を合理化

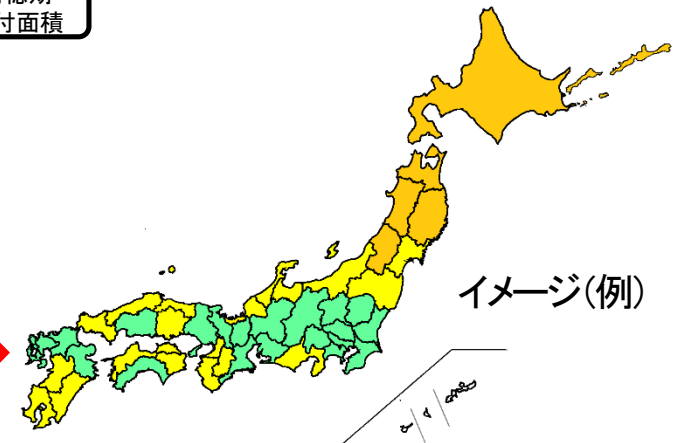
人工衛星からの取得データを活用した水稲作柄予測手法

気象データ（降水量、気温、日照時間、風速等）及び人工衛星データ（降水量、地表温度、日射量、植生指数等）を説明変数、10a当たり予想収量を目的変数として予測式（重回帰式）を作成し、水稲の作柄を予測。

【予測式の作成方法】



区分	作柄の良否	都道府県数
（オレンジ色）	やや良	5
（黄色）	平年並み	20
（緑色）	やや不良	21



人工衛星からの取得データを利用した水稲作柄予測手法を導入することにより、実測調査に係る業務を合理化しつつ、水稲作柄を予測。

SDGグローバル指標における人工衛星データの利活用 別紙

SDGsの進捗状況を測るグローバル指標について、人工衛星データを用いて算出・検証

15 陸の豊かさを
守ろう



指標15.4.2（山地グリーンカバー指数：山地総表面に対する植生被覆の割合）

ターゲット15.4の進捗をモニタリングするための指標として設定

ターゲット15.4 2030年までに持続可能な開発に不可欠な便益をもたらす山地生態系の能力を強化するため、生物多様性を含む山地生態系の保全を確実に行う。

- FAOによる試算値を、人工衛星から取得した土地被覆データを用いて検証。うち一部の分類について、FAO試算値が過大となっていた。
- 検証結果はレポートにまとめ、総務省ホームページで公表している。また、算出値は外務省ホームページで公表している。

ビッグデータ連携会議レポート（SDG15.4.2（山地グリーンカバー指数）の検証）

https://www.soumu.go.jp/main_content/000763967.pdf

JAPAN SDGs Action Platform（SDGグローバル指標：ゴール15）

<https://www.mofa.go.jp/mofaj/gaiko/oda/sdgs/statistics/goal15.html>



ESA-CCI土地被覆分類

11 住み続けられる
まちづくりを



指標11.3.1（人口増加率と土地利用率の比率）

ターゲット11.3の進捗をモニタリングするための指標として設定

ターゲット11.3 2030年までに、包摂的かつ持続可能な都市化を促進し、全ての国々の参加型、包摂的かつ持続可能な人間居住計画・管理の能力を強化する。

算出方法

人口増加率に対する土地利用率（LCRPGR）の比率は、次の数式を使用して計算される。

$$\text{人口増加率に対する土地利用率（LCRPGR）} = \frac{\text{土地利用率（LCR）}}{\text{人口増加率（PGR）}}$$

NASA及びConservation International等による指標の算出を支援するプラットフォーム



<http://trends.earth/docs/en/>

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

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 | <ul style="list-style-type: none"> ・ 携帯基地局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。 ・ 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。 ・ AIによる人工衛星データ解析、アメダスデータを利用した水稻の作柄予測。 |
| 民間ビッグデータの利活用や統計的分析事例の紹介 | <ul style="list-style-type: none"> ・ 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。 ・ 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。 |
| 民間から提供を受けたデータを用いた分析の報告 | <ul style="list-style-type: none"> ・ 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を利活用し、経済活動のナウキャストを行った検証の紹介。 ・ S社が保有するプローブカーデータを活用した交通の可視化事例の紹介。 |

構成員

高橋 大志 慶應義塾大学大学院教授（座長）
 和泉 潔 東京大学大学院教授
 大西 浩史 日本データマネジメント・コンソーシアム理事 兼 事務局長
 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

ビッグデータ連携会議の開催実績（2018～2019年度）

参考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 商業動態統計試験調査（経済産業省）	WG結果の報告。 POSデータ（大型家電量販店） を使用した 商業動態統計試験調査 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。

ビッグデータ連携会議の開催実績（2020～2021年度）

参考2

開催日	テーマ名	概要
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。
第11回 R2.12.23	労働ビッグデータの利活用（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 」の検証結果報告。
第12回 R3.3.10	位置情報による経済活動のナウキャスト（日本銀行）	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 を利活用し、 経済活動のナウキャスト を行った検証の紹介。
	観測データ利活用検証WG報告	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 最終報告 。
第13回 R3.6.30	データマネジメントの重要性（JDMC）	データマネジメント を行い、データに 付加価値 を与え、ビジネスに利活用した事例紹介
	損害保険に関するBDの活用（M社）	M社のデータを利活用し、 統合データ から データの詳細化 を行う取組紹介。
	観測データ利活用検証WG（SDG11.3.1）キックオフ	SDG11.3.1（人口増加率と土地利用率の比率） のキックオフ。
第14回 R3.9.29	オルタナティブデータ活用の黎明期を超えるために（和泉構成員）	オルタナティブデータ に横串を刺した、 AIセンシングセンター 構想の紹介
	農業統計における人工衛星データ利活用について（農林水産省）	AI による 人工衛星データ 解析、 アメダスデータ を利用した 水稻の作柄予測 。
第15回 R3.11.17	クラウド型会計ソフトから得られる決算データについて（T社）	T社の顧客約 25万社99業種 の 月次決算のデータ を集めた、月次指標（月次BAST）の紹介
	EBPM推進委員会データ利活用WGとりまとめについて（内閣官房）	データ利活用WG の取りまとめの概要紹介
第16回 R4.2.3	人流データを活用した宿泊旅行統計調査延べ宿泊者数の推計について	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 を利活用した、 宿泊者数の推計検証 の報告
第17回 R4.3.18	プローブカーデータの利活用について（S社）	S社が販売するプローブカーデータを活用した交通の可視化事例の紹介。
	観測データ利活用検証WG報告	SDG11.3.1（人口増加率と土地利用率の比率） の 最終報告