

# 「ビッグデータ連携会議」における取りまとめの 修正の方向性について

注 本資料は、次ページ以降左欄に「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）」を記載、右欄に修正の方向性と事務局におけるヒアリング結果などの関連情報を記載したもの。

その他、取りまとめに際しては、各活用事例の最新の状況やこれまでのビッグデータ連携会議の開催実績等を踏まえ、適宜時点の修正等を行う。

<p>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)</p>	<p>修正の方向性と関連情報</p>
<p><b>1 本資料の目的</b>  「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<sup>1)</sup>」(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。</p>	<p>【修正の方向性】</p> <ul style="list-style-type: none"> <li>ビッグデータ連携会議の開催実績等を追記。</li> </ul>
<p><b>2 ビッグデータの活用法やメリット</b>  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</p>	

<sup>1)</sup> ビッグデータ連携会議：[https://www.soumu.go.jp/main\\_sosiki/kenkyu/big\\_data/index.html](https://www.soumu.go.jp/main_sosiki/kenkyu/big_data/index.html)

計の作成等にも寄与することが期待されている。	
<p><b>(1) 既存統計の補完</b></p> <p>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。</p>	
<p><b>① 速報性</b></p> <p>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、CTI<sup>2</sup>）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、従来よりも速報性に優れた指標を開発中である。</p>	
<p><b>② 詳細化</b></p> <p>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</p>	

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

<p>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。</p>	
<p><b>③ カバレッジの拡大</b>  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、CPI<sup>3</sup>）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、価格取集期間や高頻度（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。</p>	
<p><b>④ 新たな指標等の作成</b>  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<sup>4</sup>などを始め、スーパーマーケット、コンビニエンス</p>	<p><b>【修正の方向性 1】</b></p> <ul style="list-style-type: none"> <li>「(1)既存統計の補完」の一部として整理していた「新たな指標等の作成」を(2)として別立てして記載。</li> </ul> <p><b>【関連 第15回ビッグデータ連携会議議事概要抜粋】</b></p> <ul style="list-style-type: none"> <li>これまで連携会議で扱った事例には、現在の公的統計を置き換えたりサポートしたりするものと、新しい公的統計を作ろうとするものがあった。既存の公的統計を置き換えるものと新たにビッグデータの特徴を活かしたものを作るという</li> </ul>

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

<sup>4</sup> SNS×AI 景況感指数、鉱工業生産予測指数：[http://qr.nomura.co.jp/quants/sns\\_ai/](http://qr.nomura.co.jp/quants/sns_ai/)

<p>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<sup>5</sup>などの開発を試みている。</p>	<p>2つの系統があることを示すと特徴が出るかと思う。</p> <p>【修正の方向性 2】</p> <ul style="list-style-type: none"> <li>地球観測衛星データを活用したSDG指標の取組を追記。</li> </ul> <p>【関連資料】</p> <ul style="list-style-type: none"> <li>連携会議レポート No01「SDG15.4.2（山地グリーンカバー指数）の検証」等</li> </ul>
<p><b>(2) 報告者負担や業務負担の軽減</b></p> <p>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<sup>6</sup>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。</p>	<p>【修正の方向性】</p> <ul style="list-style-type: none"> <li>農林水産省における「農業統計における人工衛星データの利活用にかかる取組」を追記。</li> </ul> <p>【関連資料】</p> <ul style="list-style-type: none"> <li>第14回ビッグデータ連携会議 資料3「農業統計における人工衛星データの利活用について」</li> </ul>
<p><b>3 ビッグデータの活用における課題及びその解決方法</b></p> <p>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。</p>	

<sup>5</sup> 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](https://www.meti.go.jp/statistics/bigdata-statistics/bigdata_pj_2019/pos_nomura.html)

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

### (1) ビッグデータの特性把握

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性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効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WG<sup>7</sup>)

○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)

### 【修正の方向性】

- 以下のヒアリング結果も参考にして追記を検討。

### 【関連 事務局におけるヒアリング結果抜粋】

- 景気判断での活用にあたっては、オルタナティブデータの質を検証する必要があるが、現在活用している指標については、公的統計結果等との相関にかんがみても、質が担保されており有益であると考えている。今後、新たに活用する場合は、指標の質や公的統計結果との相関など継続して検証していくことが必要。

有料データの場合は、上記の質の担保の確認が困難な場合が多いと考えられることから、サンプルとして過去のデータが見られるなどの対応があると良い。データの内訳についても開示が進むと活用が進むのではないかと。

また、データの安全性が確保された上で、民間企業によるデータ提供が進み、様々な指標が活用可能となることを期待。

(月例経済報告(内閣府))

<sup>7</sup> 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](https://www.soumu.go.jp/main_content/000630006.pdf)

<p>○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。(C T I)</p> <p>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。</p>	
<p><b>(2) ビッグデータの入手</b></p> <p>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</p>	
<p><b>① データ取扱いのルール設定</b></p> <p>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。</p> <p>○C T I 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</p>	<p><b>【修正の方向性】</b></p> <ul style="list-style-type: none"> <li>EBPM 推進委員会データ利活用 WG 取りまとめ(令和3年6月23日)にかかる取組を追記。</li> </ul> <p><b>【関連資料】</b></p> <ul style="list-style-type: none"> <li>第15回ビッグデータ連携会議 資料2「EBPM 推進委員会データ利活用 WG 取りまとめについて」</li> </ul>

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。(C T I)

- 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、I P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。(C P I)

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。

<p><b>② データ提供者へのインセンティブ付与</b></p> <p>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。</p> <p>○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)</p> <p>○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)</p> <p>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。</p>	<p><b>【修正の方向性】</b></p> <ul style="list-style-type: none"> <li>以下のヒアリング結果も参考にして追記を検討。</li> </ul> <p><b>【関連 事務局におけるヒアリング結果抜粋】</b></p> <ul style="list-style-type: none"> <li>ビッグデータは民間所有のデータであることから、購入費用の高騰リスクやデータ諸元が明らかになっていないことなどの課題がある。 (パーソントリップ (PT) 調査 (国土交通省))</li> <li>各企業において自社データの活用が広がっていると考えられ、企業側に新たな負担をかけることなく、複数企業から既存データをそのまま提供いただいているところ。したがってデータ形式は統一化されていない。また、そもそも協議会においてデータ提供は任意であり、より多くのデータをよりタイムリーに提供いただくためには、そのインセンティブ付与が課題。 (消費動向指数 (CTI) (総務省))</li> </ul>
<p><b>③ データ提供の中断に関するリスク対応</b></p> <p>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</p>	<p><b>【修正の方向性】</b></p> <ul style="list-style-type: none"> <li>以下のヒアリング結果も参考にして追記を検討。</li> </ul> <p><b>【関連 事務局におけるヒアリング結果抜粋】</b></p> <ul style="list-style-type: none"> <li>「METI×NOMURA コンシューマーセンチメント・インデッ</li> </ul>

<p>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。</p>	<p>クス」(「消費者心理指標」)については2021年6月、「SNS×AI 鉱工業生産予測指数」は2021年3月に更新を終了している。 (METI POS 小売販売額指標 [マイクロ] (経済産業省))</p> <ul style="list-style-type: none"> <li>指標の質を判断するには、継続して数値を見ていく必要がある。また、季節変動を捉えるには一定のデータ期間が必要。 (月例経済報告 (内閣府))</li> </ul>
<p><b>(3) ビッグデータを扱う人材・体制面</b></p> <p>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。</p> <p>○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。(CPI)</p> <p>○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)</p> <p>○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</p>	<p><b>【修正の方向性】</b></p> <ul style="list-style-type: none"> <li>民間事業者の活用事例等について追記。</li> </ul> <p><b>【関連 事務局におけるヒアリング結果抜粋】</b></p> <ul style="list-style-type: none"> <li>POSデータの集約、ウェブスクレイピングによる価格取集のいずれも業者に委託し、適切に業務管理を行っており、現状、特段の支障等は発生していない。 (消費者物価指数 (CPI) (総務省))</li> <li>POSデータ等の調査票への組替集計を委託。現状、特段の支障等はない。 (商業動態統計 (経済産業省))</li> <li>データの加工・集計、「BigData-STATS ダッシュボード(β版)」への掲載等は業者に委託しており、現状、特段の支障等はない。 (METI POS 小売販売額指標 [マイクロ] (経済産業省))</li> <li>データのハンドリングや分析、CTI への統合には、高度な統計手法などの知識・スキルや膨大な検証作業が必要であり、限りあるリソースの配分や人材の育成が課題。 (消費動向指数 (CTI) (総務省))</li> </ul>

<p>手引き<sup>8</sup>」を作成した。(パーソントリップ調査)</p> <p>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。</p>	
<p><b>(4) システム・技術面</b></p> <p>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。</p> <ul style="list-style-type: none"> <li>○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)</li> <li>○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ムを作成している。(C P I)</li> </ul> <p>今回の事例はデータの処理に特化したものとなってい</p>	

<sup>8</sup> 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>

<p>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。</p>	
	<p><b>【修正の方向性】</b></p> <ul style="list-style-type: none"> <li>ビッグデータを活用した指標の公表方法等に関連する事例、留意点等について記載。</li> </ul> <p><b>【関連 事務局におけるヒアリング結果抜粋】</b></p> <ul style="list-style-type: none"> <li>「BigData-STATS ダッシュボード(β版)」を経済産業省のホームページ内に開設し、試験的に指標の公表を行っている。 (METI POS 小売販売額指標 [マイクロ] (経済産業省))</li> <li>パーソントリップ調査は、必要な情報について統計的に精度担保できるよう実施し、ビッグデータは、パーソントリップ調査で把握できない人の動きを把握するために、活用した。 (パーソントリップ (PT) 調査 (国土交通省))</li> </ul>
<p><b>(5) その他 (情報保護、法令関係など)</b> 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</p>	<p><b>【修正の方向性】</b></p> <ul style="list-style-type: none"> <li>以下のヒアリング結果も参考にして追記を検討。</li> </ul> <p><b>【関連 事務局におけるヒアリング結果抜粋】</b></p> <ul style="list-style-type: none"> <li>速報性に優れたオルタナティブデータは今後も活用してい</li> </ul>

<p>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。</p> <p>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。</p>	<p>く見込み。</p> <p>(月例経済報告(内閣府))</p>
<p><b>4 活用の推進に向けて</b></p> <p>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。</p>	<p><b>【修正の方向性】</b></p> <ul style="list-style-type: none"> <li>全体の修正も踏まえて、修正。</li> </ul> <p><b>【関連 事務局におけるヒアリング結果抜粋】</b></p> <ul style="list-style-type: none"> <li>新型コロナウイルス感染症の影響などにより、CTIの改定幅が大きくなる場合があり、クレジットカード情報等による補完の可能性を研究している。</li> </ul> <p>(消費動向指数(CTI)(総務省))</p>
<p><b>別紙</b></p> <p>別紙1 ビッグデータ連携会議の概要</p> <p>別紙2 ビッグデータ連携会議の開催実績</p> <p>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</p>	