

2018年6月

民間ビッグデータを統計として活用するためには、何が必要か：
諸外国の取組事例の紹介と日本における課題の整理

別所 英実*

* 総務省 統計委員会担当室 室長補佐

総務省 統計委員会担当室
〒162-8668 東京都新宿区若松町19番1号

総務省統計委員会担当室ワーキングペーパーは、統計委員会担当室スタッフ又はスタッフと外部研究者との共同による調査・研究の成果をまとめたもので、公的統計の整備に係る各種施策に役立てることを企図としている。ただし、ワーキングペーパーの内容や意見は、執筆者個人に属し、総務省の公式見解を示すものではない。

民間ビッグデータを統計として活用するためには、何が必要か：
諸外国の取組事例の紹介と日本における課題の整理

別所 英実*

要 旨

民間ビッグデータを公的統計の「新たなデータ源」として活用するための検討が進んでいる。我が国でも、近年、政府が取り組む一連の抜本的な統計改革の中で、ビッグデータ活用の促進が大きなテーマの一つとして位置づけられている。本稿では、今後の活発な取組を促す観点から、ビッグデータ活用の本格化に向けて望ましいと思われる実務上の検討プロセスの雛形を提示した。その上で、諸外国の先進事例をサーベイしながら、ビッグデータを保有する民間主体との協力関係の構築のあり方を中心に実務上の課題について整理し、我が国への含意を探った。今後、さまざまな試行的な取組を通じて、効率的且つ効果的にビッグデータの統計的利活用を実現できるようなベストプラクティスを見出していくことが期待される。

キーワード：ビッグデータ、公的統計、統計改革

J E L分類：C55、C40

* 総務省 統計委員会担当室 室長補佐

本稿は、2018年5月23日に総務省において開催した「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」（第1回会合）における事務局報告のために行った調査・研究をベースに作成したものである。本稿の作成にあたっては、五十嵐盛仁、上田聖、櫻川幸恵、肥後雅博、松井伸司、八木雅彦（以上、総務省）の各氏から有益な助言を頂いた。ここに記して感謝したい。ただし、本稿の内容と意見は筆者個人に属し、総務省の公式見解を示すものではない。また、ありうべき誤りはすべて筆者個人に属する。

1. はじめに

デジタル技術の高度化に伴い、従来の一般的な技術では管理困難な大量のデータ群が爆発的に増加しており、こうした「ビッグデータ」の収集・分析により新たな知見を得ることを通じて、ビジネスの効率化やイノベーションに役立つ動きが拡大している。ビッグデータは、「新世紀の天然資源¹」とも言われ、我々が生きる現代社会に重大な変化をもたらしつつあるが、こうした動きは公的統計を作成する当局においても無縁ではない。近年では、統計作成当局においても、ビッグデータを統計の「新たなデータ源」として捉え、その利活用に取り組み始めている。

現状、我が国において民間企業等が保有するビッグデータ等を経常的に活用する公的統計は多くないものの、POS データやクレジットカードなど販売・購買のトラッキング情報からマクロ消費動向を捉えるための新指数の検討や、SNS 情報を人工知能（AI）によって抽出・分析の上でタイムリーな景況感指数を作成する取組など、試行的な動きもみられ始めている。

政府が取り組む抜本的な統計改革の中でも、各府省におけるビッグデータ等の新たなデータ源の積極的な活用が求められており、本年5月には総務省において、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ための「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」が立ち上がったところである。今後は、利活用上の先行事例があるデータや優先度の高いデータを具体的に選定して、産官学連携により集中的な協議を行うことを通じた利活用推進が期待されている。

しかし、我が国における検討は緒に就いたばかりであり、これから様々な統計分野における取組を積み重ねながら効率的な利活用のあり方を模索する局面にある。こうした背景を踏まえ、本稿では、統計作成当局がビッグデータ活用の検討を始めるに当たっての手掛かりとなるような検討プロセスの雛形を一つ示し、諸外国のサーベイを通じて、ビッグデータを保有する民間主体との協力関係の構築のあり方を中心に実務上の含意を引き出すことを目的としている。

¹ IBM[2013]

本稿の構成は、以下の通りである。まず、第 2 節において、ビッグデータの定義や特徴について公的統計の観点から整理した上で、公的統計への利活用上のメリットを具体的な事例を交えて紹介する。続く第 3 節では、我が国におけるこれまでの公的統計への活用状況に加えて、近年の統計改革におけるビッグデータ活用に向けた具体的な検討状況を紹介する。その上で第 4 節においては、望ましいと思われる検討プロセスの雛形を示した上で、諸外国の先行事例に学びながら実務上の含意を探る。第 5 節では、上記を要約することで結びに代える。

2. ビッグデータの概要

本節では、ビッグデータの定義や特徴について公的統計の観点から整理した上で、具体的なビッグデータの種類や公的統計への利活用上のメリットについて概観する。

2.1. ビッグデータの定義・特徴

ビッグデータを巡っては、文脈に応じて様々な解説や分析が行われており、現状では、一つの統一的な定義が確立されているわけではない。ただし、量（high-Volume）、生成速度（high-Velocity）、多様性（high-Variety）の3つの特性（3Vs）を備えたデータである点は、概ね共有された状況にある^{2,3}。

公的統計への活用を念頭にビッグデータを捉える文脈では、例えば国際連合欧州経済委員会（United Nations Economic Commission for Europe: UNECE）の議論では、各国統計作成当局が検討の射程とするビッグデータを、「3Vsの特性を備えたデータ源（data sources）」と捉えるにとどめている⁴。

こうした定義の下では、行政記録情報やスキャンデータなど個別具体的なデータを射程に含むか否かについては議論の余地を残してしまうが、ビッグデータを巡る定義上の混乱や更なる論争を避けつつも実務上の進捗を得るに十分との指摘⁵があり、特定のデータを含むか否かはビッグデータの分類方法（classification）上の論点として検討されている。

我が国においても、ビッグデータを巡る定義は一様ではない。政府が保有する行政記録情報をビッグデータに含むかどうかについても整理が区々であり、その多様性とデータ量に着目して公的統計も一

² ビッグデータが有する3つの特徴（3Vs）に、信頼性（Veracity）を加えた4Vsや、ボラティリティ（Volatility）を加えた5Vs、可変性（Variability）や価値（Value）、可視化（Visualization）などのさまざまな「Vs」を加えた指摘もみられている。

³ 当該3つのVsについては、2001年に米Gartner社（旧META Group）のDoug Laney氏が最初に指摘したと言われており、以降、ビッグデータ概念の浸透も加速したとされている。詳細は、Laney [2001]を参照のこと。

⁴ Big data are data sources that can be –generally– described as: “high volume, velocity and variety of data that demand cost-effective, innovative forms of processing for enhanced insight and decision making. (UNECE[2013])

⁵ Struijs [2016] 4頁

種のビッグデータとして捉える整理⁶や、国や地方公共団体が提供するオープンデータをビッグデータの種別の一つに含むもの⁷も存在している。本年3月に閣議決定された、平成30年度を始期とする新たな「公的統計の整備に関する基本的な計画」では、「ICTの進展により生成・収集・蓄積等が可能・容易になる多種多量の民間企業が保有するデータ⁸」と定義されており、行政記録情報はビッグデータの定義の外で別途検討されている。

こうした経緯や、行政記録情報と民間ビッグデータでは、保有主体や電子化の状況、国際比較可能性、母集団代表性の偏りに関する知見の有無など、個々に直面する課題について相違がある⁹ことを踏まえて、本稿では、上記の「公的統計の整備に関する基本的な計画」の定義に倣い、「ICTの進展により生成・収集・蓄積等が可能・容易になる多種多量の民間企業が保有するデータ」に絞って整理を試みる。

2.2. ビッグデータの種類

ビッグデータにはさまざまな種類が存在するため、統計作成当局が活用を検討する際には、検討上の射程に置く具体的なデータ範囲を事前に特定する必要がある。こうした中で、ビッグデータの種類をリスト化する試みが行われている。

もっとも代表的なものとして、UNECEが2013年に提示した分類リストが存在する。これは、ビッグデータを「生成源」(the type of generator of the data)の切り口から分類したもので、大分類として以下の3つを提示している。

1. 個人生成データ (human-sourced information)
2. 取引生成データ (process-mediated data)
3. デバイス生成データ (machine-generated data)

⁶ 會田 [2015] 2 頁

⁷ 総務省 [2017b] 53 頁

⁸ 閣議決定[2018] 4 頁

⁹ Florescu *et al.* [2014] では、調査統計・行政記録情報・ビッグデータに係る特徴について14項目から比較整理しており、行政記録情報とビッグデータについては、国際比較の可能性や従前の統計的手法の活用可能性、母集団代表性に係る偏りに関する知見の有無、データの継続性、速報性、データ取得コストなど、双方の特徴に相違がある点を指摘している。

2.2.1. 個人生成データ

個人生成データは、従来は書籍や手紙、写真などとして記録されていたような個人の生活やコミュニケーションの記録のほか、新聞や雑誌がカバーする社会の出来事の記録などがデジタル化されたものを指す（具体例は以下の通り）。

- ・ Facebook や Twitter などの SNS 情報
- ・ Instagram や YouTube などの画像や映像
- ・ Google や Yahoo!などを活用したインターネット検索記録

2.2.2. 取引生成データ

取引仲介データは、従来は取引台帳や顧客管理簿などとして記録されていた商取引に関する記録が、ネットショッピングの拡大や IT システムの発展によりデジタル化されたものを指す。

- ・ 電子商取引記録や購買履歴
- ・ 銀行決済記録や株式の売買記録
- ・ クレジットカード情報などの取引データ
- ・ 民間企業等のウェブサイト情報

なお、UNECE は、ここで挙げた財・サービス市場や労働市場、金融市場における民間のデータだけでなく、診療記録など行政サービスに係る記録も取引仲介データとして整理している。

2.2.3. デバイス生成データ

デバイス生成データは、センサーやコンピューターなど所謂 IoT デバイスの高度化に伴って測定・記録されたデータを指す。

- ・ リアルタイムカメラによる交通や天候の記録
- ・ ホームオートメーションの記録
- ・ 携帯電話位置情報や自動車走行記録などのトラッキング情報
- ・ 衛星画像
- ・ ウェブサイトのログのようなコンピューター関連のデータ

2.3. ビッグデータ活用のメリット

統計作成当局がビッグデータの活用を行うことができれば、どのようなメリットがあるだろうか。ここでは、Struijs[2016]が示す7つの利活用方法及び潜在的効果を参考に、先行事例を紹介しながら掘り下げて説明する。

1. 新たな統計の作成への活用
2. 公的統計よりも詳細な情報提供への活用
3. 公的統計の早期化への活用
4. 公的統計のノウ・キャスティングへの活用
5. 公的統計の精度向上への活用
6. 公的統計の報告者軽減への活用
7. 公的統計の作成コストの削減と効率性向上への活用

2.3.1. 新たな統計の作成への活用

「既存の公的統計が存在しない場合」には、公的統計が捕捉できないような現象把握を目的に、ビッグデータを活用した新たな統計を作成することができる。

例えば、SNS上の人々の「つぶやき」を集めることで、人々のセンチメントの計測に活用する方法が考えられる。類似の具体例として、ドイツ連銀が、「預金保険 (deposit insurance)」を検索ワードとしてGoogle検索件数の指数化に取り組んでおり、預金取り付け (bank run) を事前に予測する研究¹⁰が行われている。

日本でも、総務省が大手携帯キャリアの保有するモバイルビッグデータを活用することで、公的統計では把握できない特定地域における時間別の人流からテレワーク・デイの実施効果を定量的に検証した事例¹¹がある。

こうした事例のように、民間ビッグデータから新たな統計的分析を試みる事例のほか、既存の公的統計とビッグデータを組み合わせることで、指標を作成することも活用方法の一つとして想定される。

¹⁰ Weber, Fecht & Thum[2017]

¹¹ 総務省[2017c]

2.3.2. 公的統計よりも詳細な情報提供への活用

「公的統計が存在する場合でも、公的統計より詳細な情報」をビッグデータから抽出することも可能である。例えば、既存の公的統計では、サンプル数の制約などから地域レベルでの統計結果が抽出できない場合や、月次ではなく四半期統計しか存在しない場合などは、ビッグデータを活用することで公的統計から詳細性を引き出すことが可能となる場合がある。

Struijs[2016]では、一例として Google Trends から地域単位でのトレンドを捕捉することで公的統計から地域単位での動向を抽出できる可能性に触れている。同時に、こうした活用方法は想定される以上に利活用が困難であることを指摘している点には留意が必要である。

2.3.3. 公的統計の早期化への活用

「公的統計の公表早期化を企図して、既存のデータ源をビッグデータで代替する」ことも可能である。質の高い統計を作成することと統計を早期に作成することは、いずれも統計作成当局にとって大きな目標であるが、統計調査や行政記録情報を活用して統計を作成する際には、実務上相応の時間が不可避である。こうした統計作成上のデータ源をビッグデータで代替することができれば、よりタイムリーな統計作成が可能となる。

この点、諸外国では、POS データを物価算出に活用する取組が複数みられている。また、エストニア中銀の取組事例では、国境に係る公的統計の活用に加えて、携帯電話の SIM カード情報を活用することで旅行客の消費を捕捉して旅行収支を推計している¹²。

このほかにも、ビッグデータと他のデータとの相関性を見出して推計モデルを構築することで、よりタイムリーな統計公表のために活用することも想定される。

2.3.4. 公的統計のナウ・キャストへの活用

「公的統計をベンチマークとしてビッグデータからナウキャスト指標を作成する」ことができる。この場合、既存の公的統計が最終的

¹² 当該事例の場合には、政府の予算制約の都合で、従来活用していた国境に係る調査が中止になったことが取組の契機となった。詳細は、Kroon[2012]、Hammer *et al.* [2017] 18 頁を参照のこと。

に公表されるため、ビッグデータを活用した指標の質が多少劣る場合でも、より早い動向把握を求める統計ユーザーのニーズに応えるに十分な新指標を開発できる場合がある。公的統計の参考指標として広く社会に役立つ可能性も期待される。

こうした活用方法は、これまでも幅広い統計分野で、さまざまなデータを活用しながら実施されている。これは、統計作成当局でなくとも統計ユーザー自らが実施することが容易であるためと思われる。先行事例も相応に蓄積されている利用法であるほか、ユーザーの関心が高い利活用方法であることから、ここでは先行事例を幾分手厚めに紹介しておく。

米 MIT の研究者が中心となって行う「ビリオン・プライス・プロジェクト」では、オンライン市場における日次データを活用することで、2008 年のリーマン・ブラザーズの破綻に際して小売り各社が殆ど直ちに販売価格を引き下げる動きを捕捉した。公的統計の公表前にタイムリーに動向を把握した事例¹³として注目されている。

このほか、人工衛星が撮影する衛星画像から夜間の光量データを把握することで、GDP のナウ・キャスティングを試みる研究¹⁴や、宿泊施設が公表しているウェブ予約のデータの分析を通じて、日々の空室数・客室稼働率を算出する研究¹⁵が存在する。

経常的な統計作成を企図せずとも、震災や停電など稀に発生するイベント前後での景気変動の把握にも有効である。例えば、デビットカード決済高を活用することで 9.11 テロや SARS などのイベントによる消費動向への影響を分析した事例¹⁶がある。我が国でも、平成 26 年の消費税率引上げによる経済への影響を具に把握するために、内閣府が POS データを週次分析した事例¹⁷などがある。

2.3.5. 公的統計の精度向上への活用

「公的統計の精度向上」を目的として、公的統計の調査結果の整合性や正確性を検証するためにビッグデータを活用することもできる。例えば、ビッグデータから捉えられる現象と公的統計の結果をクロ

¹³ Cavallo & Rigobon[2016]

¹⁴ Henderson, Storeygard & Weil[2012]

¹⁵ 一藤・曾根原[2016]

¹⁶ Galbraith & Tkacz[2013] (131-133 頁)

¹⁷ 内閣府[2014]

スチェックすることで、公的統計のクオリティチェックを実施することや、公的統計の作成に用いる推計方法の見直しなどに役立てることが想定される。英国では、年央における人口推定の精度向上を目的に、位置情報付き Twitter の情報活用を検討した事例¹⁸がある。

また、時代に相応しい正確な調査統計を提供するためには、調査事項や調査項目、調査方法などを事前に設計しなければならないが、合理的且つ妥当な調査となるよう企画や実査の段階でビッグデータを活用することも考えられる。

例えば、英国では、不動産仲介会社のサイトの掲載情報からトレーラーハウスの情報を収集し、アドレス・レジスターの登録情報の拡大を試みた事例¹⁹がある。このほか、Bean[2016]は、ネット求人情報の活用を念頭に、新たに登場しつつある業種や職種分類などの整備にビッグデータ活用が有効であると指摘している。

2.3.6. 公的統計の報告者軽減への活用

「統計調査をビッグデータで代替することで、報告者負担を軽減する」ことができる。例えば、調査票の記入情報から作成する物価指数について、インターネット上の物価情報で代替する事例が存在する。

既存の公的統計を「完全には代替できない場合」でも、公的統計をビッグデータで補完することができれば、統計調査におけるサンプル数や調査頻度、調査項目等の削減が達成される場合がある。

2.3.7. 公的統計の作成コストの削減と効率性向上への活用

報告者負担の軽減を図るようなビッグデータの活用は、「統計作成当局にとっての負担軽減と業務効率の向上」につながるケースが多い。しかし、統計作成当局は両者を別個の目標として設定するため、当局自身の業務効率化を起点として、ビッグデータの収集を図ることが想定される。追加的なコストや業務負担を回避しつつ、利用可能なデータを最大化させことできれば、作成コストの削減と効率性向上が達成される。

¹⁸ 年央の人口変化の原因はさまざま、就学や就職に伴う引っ越しなどが想定されるが、各地域の公営診療所への登録情報のみをベースに推計を実施すると、引っ越し後も登録を先送りする傾向の高い層の人口推計が困難となるため。詳細は、Sweier, Komarniczky & Clapperton[2015] を参照のこと。

¹⁹ ONS[2017]

2.4. 統計作成当局のスタンス変化

このようなビッグデータのメリットを示す事例が蓄積するにつれて、統計作成当局は、ビッグデータは、もはやパスワードではなくなりつつある²⁰との認識の下で、具体的な戦略や取組に乗り出している。各国統計局のみならず国際連合や国際通貨基金などの国際機関がグローバルに取組を後押しする中で、ビッグデータは統計作成当局にとっての「ゲームチェンジャー²¹」とみる向きも存在する（国際的な取組の概観は補論を参照）。

こうした取組が、民間ビジネスにおけるビッグデータ活用に比べて幾分遅行するかたちで活発化した背景について若干考察しておく。 (イ) 従来の公的統計の作成手法で十分対応できたこと、(ロ) ビッグデータの利活用上の不透明感の存在、(ハ) 公的統計が築いてきた信頼や評判を失うリスクの存在²²、等が影響したものと考えられる。

統計作成当局によるビッグデータ活用を巡る「リスク評価」は、デジタル経済の進展に伴う利用可能なビッグデータの増大や、先行的な優良事例が蓄積する中で変化してきた。民間ビジネスにおいてもビッグデータを活用した民間統計が多数生成されるような時代を迎える中で、統計作成当局による統計作成に係る独占的地位が揺らいでいる²³。統計作成当局がビッグデータ活用の検討に踏み込まないことに伴うリスクと、活用に乗出すことに伴うリスクのバランス変化が、統計作成当局の時代認識の変化²⁴に繋がったものと思料される。

もっとも、統計調査を巡る環境や報告者のスタンス、公的統計の質を巡る課題など、各国の統計作成当局が直面する環境は国情に応じてさまざまであり、従ってビッグデータの活用に向けたスタンス及びその必要性が一様ではない点は、各国の動向をみる上で留意すべきポイントの一つである。

²⁰ Hammer *et al.* [2017] 29 頁

²¹ Struijs[2016] 3 頁

²² Bender[2017]

²³ Eurostat [2017] 3 頁

²⁴ 我が国でも、西村清彦統計委員会委員長が、ビッグデータ活用に向けた検討は避けられないとの時代認識を示している（総務省統計委員会[2017a]）。Eurostat [2017]では、統計作成当局が情報マーケットから疎外されるリスクを指摘するなど、ビッグデータを活用しないことへの大きな危機感が窺われる。

3. 日本における議論・取組状況

国際的にビッグデータ活用の検討が進む中、本節においては、我が国における利活用の状況や、各府省の取組について確認する。また、近年政府が取り組む抜本的な統計改革の文脈から、ビッグデータの積極的な活用に向けた検討状況を紹介する。

3.1. 各府省による取組状況

現状、我が国において民間企業等が保有するビッグデータ等を「経常的に」活用している公的統計（図表 1）は限定的であり、そのうち本稿が念頭に置く「ICT の進展により生成・収集・蓄積等が可能・容易になる多種多量の民間企業が保有するデータ」を活用した事例は、消費者物価指数への POS データ活用に留まっている。

図表 1：民間企業等が保有するビッグデータ等を経常的に活用している統計

統計名 実施機関		民間データの名称	民間データ 収集の根拠	活用形態	活用による効果
国民経済計算 (加工統計)	内閣府	有価証券報告書等	原則、オープンなデータを取得している	各種基礎データのの一つとして活用	他の統計で得られない国民経済計算に必要なデータを得られる
景気動向指数 (加工統計)	内閣府	<ul style="list-style-type: none"> ・日経商品指数（42種総合） （株）日本経済新聞社） ・東証株価指数 （株）東京証券取引所） ・長期国債（10年）新発債流通利回り （日本相互証券（株）） ・中小企業売上げ見通しDI （日本政策金融公庫） （企業物価指数（日本銀行）） （マネーストック（M2）（日本銀行）） 	オープンなデータを取得しており、協定等は締結していない	景気動向指数・先行指数の11の採用系列のうち6系列に活用	景気動向指数・先行指数は景気に対し先行して動く指標の動きを統合している。マーケットの動きや、消費者・企業のマインドを測るデータは先行性があり、これらのデータを使うことで景気動向の把握をより適切に行うことができる
消費者物価指数 (CPI)	総務省	POSデータ	契約ベース（POSデータ保有企業1社から平成12年以降、月次で購入している。）	「パソコン（デスクトップ型）」、「パソコン（ノート型）」、「カメラ」の3品目については、当該データのみによって指数を作成	品質向上が著しく製品サイクルが極めて短いパソコン及びカメラについて、品質調整済みの価格変動をヘドニック法により直接求めており、POS情報による販売価格、販売台数、各機種の特徴などを用いることにより、客観的で信頼度の高い重回帰分析を行うことが可能となる
設備工事業に係る受注高調査 (加工統計)	国土交通省	<ul style="list-style-type: none"> ・（一社）日本電設工業協会 ・（一社）日本空調衛生工事業協会 ・（一社）日本計装工業会の独自集計資料 	協力ベース（依頼に基づき、平成3年以降、データの提供を受けている。）	当該データのみによって統計を作成	調査を行わず統計の作成が可能となる

（注）ビッグデータを含む民間が保有するデータ全般広く含む整理。

（資料）総務省ホームページをもとに作成。

他方で、総務省による POS データやクレジットカードなどの販売・購買のトラッキング情報からマクロ消費動向を捉えるための新指数の検討²⁵や、経済産業省による SNS 情報を人工知能（AI）によって抽出・分析の上でタイムリーな景況感指数を作成する取組²⁶など、各府省主導の「試行的な取組」も始まっている。

経済産業省による取組については、家電量販店の販売動向や人々の景況感を捕捉するための指数、鉱工業生産指数を予測するための指数がすでに試験公表されており、ユーザーからの意見を求める段階を経て次の段階への移行を検討中である。また、消費者物価指数についても、インターネット通信販売価格や宿泊料について、ウェブスクレイピングによるデータ収集や分析の取組²⁷が行われている。

経常的な統計作成への活用を目指したもの以外にも、インターネット上の記事から消費者マインドを指数化する取組²⁸など、景気判断を行う際にスポット的に活用する事例がある。また、地方創生に向けた取組を情報面から支援するために提供されている「地域経済分析システム」（RESAS）では、公的統計だけでなく民間企業が保有するさまざまなビッグデータが活用されており、行政のみならず、民間企業による地域マーケティングなどにも利用されている。

上記で示した政府の取組例のように、各府省による積極的な利活用が拡大しつつあるが、近年の統計改革の中で、更なる利活用推進が期待される状況にあり、これを後押しする具体的な仕組みの設計が進捗している。

3.2. 産業界からの提言

公的統計は、行政施策の企画・立案・評価のための基礎データとし

²⁵ 総務省は、2017年7月に、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することを目的として、消費動向指数研究協議会を設立。詳細は、総務省[2017d]を参照のこと。

²⁶ 経済産業省は、2017年7月に、新たに「BigData-STATS」と呼称されるウェブサイトを設け、SNS×AI 景況感指数、SNS×AI 鉱工業生産予測指数、POS 家電量販店動向指数を開発・試験公表。詳細は、経済産業省[2017a]、同[2017b]を参照のこと。

²⁷ 詳細は、総務省[2017a]、同[2018a]を参照のこと。

²⁸ 内閣府[2018]では、ニュース記事データを用いて消費者マインドと相関が高いと思われる指標を作成。分析の結果、消費者マインドの変化と概ね似通った方向感で推移しているものの、同指標とマインドの動きが異なる部分もみられたことから、速報性の高い情報として活用する際には更なる改善が必要である、と指摘している。

て活用されるのみならず、個々の世帯や民間企業が合理的に意思決定を行うための基盤である。民間企業は、公的統計における調査対象としての報告者であり、且つ統計ユーザーとしての的確な経営判断を行うために公的統計を活用している。かかる観点で、産業界はユーザー及び報告者双方の立場から公的統計の整備に関して積極的に関与しており、さらには、統計改善に関する具体的施策の要望や提言などを通じて統計行政の発展に大きく貢献している。

民間企業が保有するビッグデータが、報告者負担の軽減と公的統計の精度向上の双方に貢献する可能性に着目し、日本経済団体連合会や経済同友会から、更なる利活用の促進が提言されている。

日本経済団体連合会が、2016年に公表した「公的統計の改善に向けた提言」では、ビッグデータ活用のメリットと課題について整理の上で、「試験的な取組を積み重ねつつ、ビッグデータの活用に関わる課題への対処を検討していくことが必要である²⁹⁾」と論じている。

経済同友会が2016年に公表した提言書では、「民間の保有するビッグデータを活用することで、政府が作成する経済統計を補完できるうえ、経済社会の実態を低コストかつ高頻度で、より早く、より詳細に把握できる可能性もあり、この面で民間企業が貢献する余地は大きい³⁰⁾」と記述している。同時に、活用可能性の高いと思われる統計分野と具体的なデータを例示しており、ビッグデータ活用の検討を本格化させる上で、貴重な手掛かりを示している。

図表2：経済統計へのビッグデータの活用可能性・活用例

分野	利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 ⇒ 商品別・購入者属性別物価 ・ウェブサイトからの価格情報の自動収集（スクレイピング） ・不動産価格（ソニー不動産、リクルートなど）
労働	・求人情報（リクルートなど）
消費	・POSデータ ⇒ 商品別購入数量・金額 ・ポイントカード、クレジットカード ⇒ 購入者属性別・商品別購入額 ・家計簿ソフト ⇒ 購入者属性別・品目別購入額
電力	・スマートメーター ⇒ 30分ごとの全世帯・企業の電力使用量
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

（資料）経済同友会[2016]をもとに作成。

²⁹⁾ 日本経済団体連合会[2016] 11 頁

³⁰⁾ 経済同友会[2016] 7 頁

3.3. 政府における統計改革

2015年10月の経済財政諮問会議における、基礎統計の充実に関する麻生財務大臣の問題提起を契機として始まった近年の統計改革でも、ビッグデータ活用の推進は、大きなテーマの一つと位置づけられ、利活用促進のための具体的な仕組みづくりが進展している。

2016年末の経済財政諮問会議で取り纏められた「統計改革の基本方針³¹⁾」では、早期かつ精緻な景気動向把握に資するビッグデータの活用推進や、ビッグデータ活用に関する環境・体制整備が提言された。さらには、抜本的な統計改革及び一体的な統計システムの整備等を政府が一丸となって強力に推進するために設けられた統計改革推進会議（議長：内閣官房長官）においても、ビッグデータの活用推進に向けた具体的な検討が進捗した。

その結果として、同年5月に統計改革推進会議が決定した「最終とりまとめ³²⁾」では、民間企業等が保有するビッグデータ等について、利活用上の問題を集中的に解決するパイロット的な枠組みの構築など具体的な措置に踏み込んだ方針が示され、これらの内容を盛り込んだ「公的統計の整備に関する基本的な計画」（以下、「第Ⅲ期基本計画」と言う。）が本年3月に閣議決定するに至った。

下図は、第Ⅲ期基本計画の別表³³⁾に記載されている「今後5年間に講ずる具体的施策」のうち、ビッグデータ関連の施策を抜粋したものである。ビッグデータの利活用の各種課題を集中的に解決するための産官学連携による会議の開催のほか、景気動向把握の向上に資するビッグデータを活用した経済指標の開発など幅広い施策メニューが並んでいる。別表には記載されていないが、基本計画本文では、統計委員会が果たすべき役割の一つとして「行政記録情報や地方公共団体・民間が保有する各種データの統計的利活用について、技術的・中立的観点から支援する」と記載されるなど、一連の統計改革の中で、ビッグデータ活用に向けた期待がいかに大きいかが示唆される。

³¹⁾ 経済財政諮問会議[2016] 平成28年12月21日決定

http://www5.cao.go.jp/keizai-shimon/kaigi/minutes/2016/1221_2/shiryo_04.pdf

³²⁾ 統計改革推進会議[2017] 平成29年5月19日決定

https://www.kantei.go.jp/jp/singi/toukeikaikaku/pdf/saishu_honbun.pdf

³³⁾ 第Ⅲ期基本計画は、第Ⅱ期基本計画までの基本計画の本文及び別表形式を踏襲し、本文に取組の経緯や必要性、今後の方向性、継続的な取組事項等を、別表に今後5年間に講ずる具体的な措置・方策、担当府省等を記載している。

図表 3：第Ⅲ期基本計画（今後 5 年間に講ずる具体的施策）

具体的な措置・方策等	担当府省
<p>POSデータ、人流データ、SNSデータなど民間データ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。）を選定して、各府省の協力により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ための産官学連携による会議を開催し、データの保護や取得等の状況にも留意しつつ、各府省と地方公共団体・民間企業等の間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を推進する。</p>	<p>総務省、各府省</p>
<p>統計的分析や統計作成目的によるビッグデータ等の利活用を推進するため、各府省におけるビッグデータ等の統計への活用実態を把握するとともに、可能な限り地方公共団体・民間企業等における国の統計データやビッグデータ等の効果的な利用状況の把握に努め、定期的にこれらの情報を各府省に提供することで各府省による利活用の横展開を促すとともに、上記の会議に報告する。</p>	<p>総務省、各府省</p>
<p>ビッグデータを用いた新たな景気動向の把握のため、POSデータをきめ細かく分析に利用する手法の開発に向けた検討を行う。また、物流データを活用した地域間の移出入の動向把握に向けて、調査機関と連携して研究を進める。</p>	<p>内閣府</p>
<p>ビッグデータを活用した経済指標の開発に当たっては、景気動向把握の向上に資するよう考慮するとともに、POSデータ、人流データ、SNSデータ等を用いて、既存統計で把握できていない経済活動の把握に努める。また、各府省におけるビッグデータの効率的な活用を推進するため、関係府省の取組状況や企業等からのデータ提供の在り方、データの品質確保、専門人材の育成等について、統計委員会を中心に定期的な情報共有を図る。</p>	<p>総務省、関係府省</p>
<p>ビッグデータの特性把握や偏り是正等ビッグデータの利用に関する高度な統計技術の研究開発に引き続き取り組む。また、匿名化などの適切なデータの補正、データ保管・利用に関する信頼性の確保、データの品質確保などについて、各府省に対する技術的な支援の充実に努めるとともに、ビッグデータの分析事例や活用事例に関する研修を実施する。</p>	<p>総務省</p>

（注）実施時期については、いずれも平成 30 年度を始期とする。
（資料）閣議決定[2018]をもとに作成。

3.4. 産官学協議のための連携会議

閣議決定された第Ⅲ期基本計画を踏まえて総務省は、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ための「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」を本年 5 月に開始した。

本会議は、検討事項として、（１）官民における、統計的分析や統計作成におけるビッグデータ等の利活用の先行事例及び先行研究の分析について、（２）統計的分析や統計作成における優先度の高いビ

ッグデータ等の選定と応用可能性について、(3) 関係者との情報共有及び優良事例の横展開の可能性について、(4) データ保護及びデータ取得方法について、等を掲げている。また、本会議で選定されたデータやテーマについて、個別にワーキンググループを設けて集中的に検討することも想定³⁴されており、産官学における専門家や関係者の連携深化を図りながら、ビッグデータの統計的利活用が一層進展することが期待される。

³⁴ 総務省[2018c]

4. 民間ビッグデータを活用した統計作成への課題

そもそも、公的統計とビッグデータは性質が大きく異なるものである。統計調査が統計作成を目的として予め調査設計されたデータであるのに対して、ビッグデータの場合には、一般の経済社会活動の副産物として生成・蓄積されたデータ³⁵と整理することができる。

ここで改めて公的統計の作成プロセスを確認すると、一部の悉皆調査を除く多くが、標本調査によって母集団全体の推計を行っており、統計学の専門的な知見を活用するかたちで、予め調査の目的や調査事項、調査対象、時期、調査方法を企画・設計している。我が国においては、各府省による統計調査が、合理的かつ妥当な設計か否か、他の調査と過度な重複がないかについて統計法の規定により総務大臣の審査・承認を受けることになっているほか、個別の統計調査に関しては統計委員会においても議論がなされている。

このように、統計作成当局は、統計作成にあたって品質コントロールを左右できる立場であることが前提となる。しかし、民間が保有するビッグデータをデータ源として活用する場合には、データの調査主体（data producers）ではなく、むしろ消費者（data consumers）として向き合わざるを得ない³⁶ところに難しさがある。ビッグデータを公的統計のデータ源として活用するためには、特定の目的や理論に基づいた演繹的な手法ではなく、帰納法的に有益なデータを抽出する発想に基づく³⁷ため、統計作成プロセスのあり方を根本から考える必要がある。

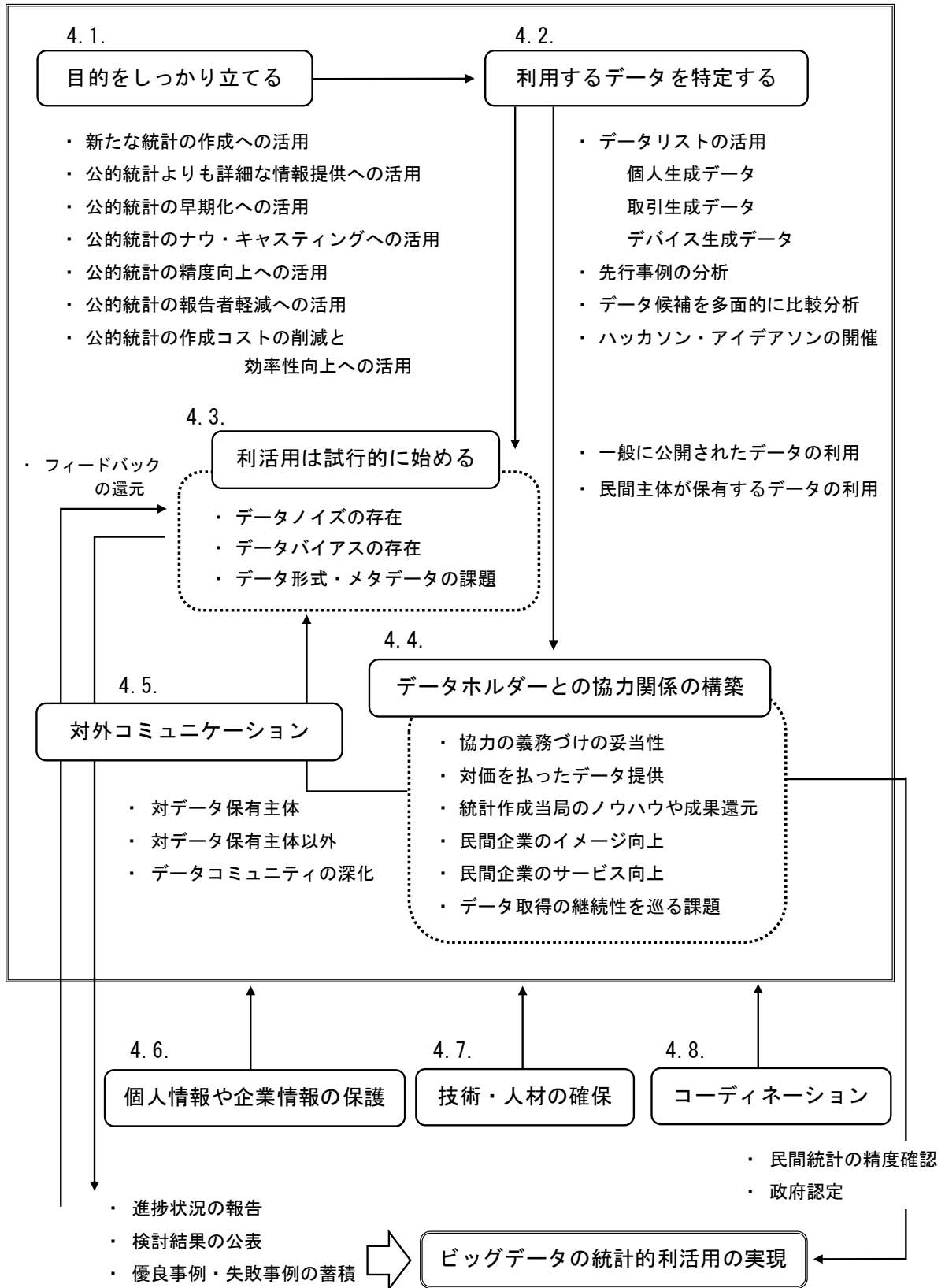
こうした中で、我が国におけるビッグデータ活用に向けた検討は緒に就いたばかりであり、実務上の検討プロセスも定まっていない状況にある。そこで本稿では、将来の活発な取組を促す観点から、望ましいと考えられる検討プロセスの雛形（図表 4）を一つ提示する。そのうえで、諸外国の先行事例に学びながら、我が国においてビッグデータ活用を本格化させるにあたっての実務上の含意を探りたい。

³⁵ Hammer *et al.* [2017]

³⁶ Scannapieco, Virgillito & Zardetto[2013]

³⁷ 北村[2018] 2頁

図表 4：ビッグデータ活用に向けた検討プロセス



(資料) 筆者作成

4.1. ビッグデータ活用の目的をしっかりと立てる

ビッグデータ活用の検討を開始するにあたっては、最初のステップとして、「取組の目的をしっかりと立てる」ことが肝要である。

第2節で紹介したようにビッグデータ活用の潜在的効果は多岐に及ぶが、活用方法次第で特定のデータが複数の効果を有することが少なくない。そのため、まずは目的をしっかりと定めた上で、目的達成に相応しいデータを選定し、目的達成に必要な範囲内で取組を行うことが効率的な検討を可能とする。

この点、既に先行する諸外国の取組においても、実際にデータ収集に取りかかる前に、取組の目的をしっかりと定めた上で検討対象となるデータを特定している。例えば、イギリス国家統計局が2014年に開始したプロジェクトでは、居住状況の調査、人流の把握、物価動向の把握、人口推計といった利活用上の目的を定めつつ、検討するデータとしてスマートメーター、モバイル位置情報、インターネット上の物価情報、位置情報付 Twitter 情報をそれぞれ選定している。

図表5：英国国家統計局によるパイロット・プロジェクト

目的	データ源
各住所における居住状況の調査	スマートメーター
人流の把握	モバイル位置情報
物価動向の把握	インターネット上の物価情報
人口推計	位置情報付 Twitter 情報

(注) 上記の目的のほかにも、技術的な側面等にも着目して具体的な目的を複数定めてプロジェクトに取り組んでいる。

(資料) Naylor *et al.*[2015]をもとに作成。

このほか、国際通貨基金が2015年に立ち上げたプロジェクト³⁸でも、国際資金フローを把握するためのSWIFTデータ³⁹の活用や、GDPナウ・キャスティングのためのGoogle Trendsの活用などのように、目的をしっかりと定めた上で、活用するビッグデータを選定している。

³⁸ 国際通貨基金は、2015年11月にプロジェクト（Big Data Innovation Challenge）を立ち上げており、6つの先駆的なアイデアの実証実験（proof of concept development）に取り組んでいる。詳細は補論を参照のこと。

³⁹ SWIFT（Society for Worldwide Interbank Financial Telecommunication）は、銀行間の国際金融取引に係る事務処理の機械化、合理化および自動処理化を推進するため、参加銀行間の国際金融取引に関するメッセージをコンピューターと通信回線を利用して伝送するネットワークシステムのこと。

4.2. 利用するデータを特定する

目的設定に続くプロセスとして、当該目的達成に「最も相応しいビッグデータを特定する」必要がある。

4.2.1. データリストの活用

統計作成当局が射程に捉えるビッグデータは、個人生成データ、取引生成データ、デバイス生成データを含めて相応の数にのぼるため、まずは、検討の対象となるデータ候補を絞り込む必要がある。その際には、第2章で紹介した UNECE リストのような既存のデータリストを検討の手掛かりとして活用することができる。

一例として、IMF スタッフによる分析では、UNECE が示したリストをもとに、行政記録情報、法人ホームページ、オンラインニュースなどのサブカテゴリーを独立して新規に設ける一方、動画データや診療情報、画像データなどは統計作成目的に活用するには時期尚早としてリストから削除して改訂版⁴⁰を示している。このように、既存のビッグデータリストをベースに、目的に応じて個別のデータをアドホックに追加ないし削除することで、候補となるデータを効率的にリスト化できる。

4.2.2. 先行事例の分析

リストからデータを絞り込む一つの方法として、内外における先行事例の分析が有効である。例えば、「労働市場における雇用動向を公的統計よりもタイムリー且つ詳細に把握する」といった取組目標を定めた場合、先行事例をサーベイすることを通じて、SNS 情報、インターネット検索情報、商取引記録、銀行決済記録、民間企業ウェブサイト（求人広告）程度まで利用可能性の高いデータ源を絞り込むことができる⁴¹。

4.2.3. データ候補を多面的に比較分析

ただし、先行する優良事例がある場合でも、データごとの活用可能性は、各国の統計整備の状況やデータ環境に大きく依存する（actual

⁴⁰ Hammer *et al.* [2017] 9 頁

⁴¹ 労働市場の民間ビッグデータの活用事例は、高田・別所・五十嵐[2018]を参照のこと。

opportunities for individual countries may be asymmetric⁴²⁾ 点には留意が必要である。従って、データ候補の絞り込みに当たっては、国情を踏まえた多面的な比較検討もまた重要となる。

ビッグデータの生成源に着目した分類で捉えられない観点としては、①データアクセスのハードル、②データアクセスの継続性、③データのプライバシー性（法的制約など）、④コスト（データ取得に関するもののほか、データマネジメント上の費用を含む）、⑤データの発展段階⁴³⁾などが挙げられる。

こうした多面的な観点からデータの利用可能性を捉えるためのデータリストを作成する取組も行われている⁴⁴⁾が、容易ではない。多面的な比較検討は、今後の実務上の課題の一つと言える。とりあえずここでは、エストニア中銀が携帯電話位置情報を活用した旅行収支の推計の検討開始時に行った、複数のデータ候補についての比較分析を紹介しておく。

図表 6：データ源を選定するための比較検討（エストニア中銀の例）

データ源	長所 (Pros)	短所 (Cos)
税関調査の継続 (調査主体の変更)	従前と同様の方法が継続(経験値) / 他調査との一部集約の可能性	調査コストが過大 / カバレッジやサンプリングなど統計精度に係る信頼性が不十分
宿泊情報	月次調査 / 実施が容易 / 地域間の偏りが無い / 低コスト	推計のための追加調査が必要 / アウトバウンド旅行を含まない / 推計誤差が大きい
道路センサー	データが高頻度 / 港湾や空港調査を補完する	高コスト / 推計のための追加調査が必要・計測誤差が大きい
クレジットカード情報	データが高頻度 / 消費額の推計が容易 / データ提供者が国内事業者のみ / 低コスト	推計のための追加調査が必要 / 先行する失敗事例の存在 / データノイズの存在
携帯電話位置情報	データが高頻度 / 母集団代表性が高い / 地域単位でのデータ精度 / 低コスト	取組経験が無い / 携帯キャリアとの協力関係が曖昧 / 推計のための追加調査が必要 / IT 技術の必要

(資料) Kroon[2012]をもとに作成。

⁴²⁾ Hammer *et al.* [2017] 27 頁

⁴³⁾ Eurostat [2017] (21 頁) では、ガートナー社が提唱する「ハイブ・サイクル」(Gartner's hype cycle) 論を引き合いに、ビッグデータそれぞれの発展サイクルが一樣でない点を強調。当該理論では、新技術の登場によって生じる過度の興奮や誇張(ハイブ)の後に幻滅期が到来し、それらを経た後ようやく安定期(Plateau of Productivity)に到達するとの見方。

⁴⁴⁾ United Nation[2015] 5 頁

4.2.4. ハッカソン・アイデアソンの開催

上記とは異なるアプローチとして、特定の政策課題を巡るハッカソン⁴⁵やアイデアソン⁴⁶を開催する方法も有効な手段となる。例えば、Eurostat および欧州委員会が 2017 年 3 月に開催したビッグデータ活用に係るハッカソン⁴⁷では、事前に「地域レベルにおける雇用のミスマッチ解消に資するビッグデータの活用法」と題する政策課題が設けられ、参加する各国統計局が解決法を提示する仕組み⁴⁸となっている。このほか、世界銀行が、個人、学界、産業界、非営利団体などを対象に、気候変動問題の解決につながるようなビッグデータ活用のアイデア募集を行った先行事例⁴⁹も存在する。

⁴⁵ 事前に示された課題やテーマについて、多様な参加者がチームを編成して技術や知見を持ち寄り、短期間にシステムやアプリケーションなどの共同開発を競うイベント。主に IT 系のテーマを中心に技術開発やオープンイノベーションを目的に開催されていたが、近年では設定されるテーマが拡大している。ハック (Hack) とマラソン (Marathon) を掛け合わせた造語。

⁴⁶ 事前に示された課題やテーマについて、多様な参加者がアイデアを持ち寄るイベント。ハッカソン同様に、チームを編成の上で集中的に成果を競うもので、ハッカソンの事前準備として開催されるケースが多い。アイデア (Idea) とマラソン (Marathon) を掛け合わせた造語。

⁴⁷ https://ec.europa.eu/eurostat/cros/content/what-purpose_en

⁴⁸ 欧州委員会は、ハッカソンの開催目的として、欧州における優秀なデータサイエンティストの発掘や、産業界や学界のコミュニティにおけるパートナーシップの促進なども掲げている。

⁴⁹ 世界銀行による Big Data Innovation Challenge のこと。詳細は、以下を参照のこと。
<http://www.worldbank.org/en/news/feature/2017/03/27/and-the-winners-of-the-big-data-innovation-challenge-are>

4.3. 利活用は試行的に始める

ビッグデータの活用目的とデータを特定したとしても、実際に目的を達成できるかは別の話である。ビッグデータは、公的統計と比較して多くのノイズを含んでおり、且つバイアスの大きいデータであるため、これらを克服するためのデータマネジメントが課題となる。

4.3.1. ノイズの存在

ビッグデータは、「ノイジーなデータ」と形容されることがある。Twitter 情報から消費者マインドの計測を試みたオランダの事例では、取得したツイートの半数が意味の無いおしゃべり（pointless babble）であったと指摘⁵⁰されており、いかにビッグデータが多くのノイズを含んでいるかを例証している。

多くのノイズを含むビッグデータの活用について、例えば、北村[2018]は、「鉱山で金やダイヤモンドを探し当てるようなもので、得られるごく少数の希少鉱物を探し出す効率的な手法がいまだに確立されていないように、ビッグデータのマイニングにも決定的な方法はまだない」と分析しており、データノイズがビッグデータの活用上いかに大きな阻害要因になるかが窺われる。

4.3.2. バイアスの存在

データバイアス（母集団代表性の欠如）も利活用を阻む検討課題⁵¹となる。バイアスの存在を示す一例として、我が国における SNS 利用率の調査⁵²を取り上げると、Twitter は 10～20 代に利用者が偏っており、Instagram は女性ユーザーが多い状況が確認されている。同一の世代においてさえ、SNS の媒体によってユーザーの趣味や性格などの傾向が異なるとの指摘もある。

SNS 情報以外にも、民間企業が保有するデータ提供を受ける場合には、当該企業の業界における地域間シェアや強みとする顧客の属性（平均年収等）が同業他社と比べて大きく異なる点なども考慮する必要がある。

⁵⁰ Daas & Puts[2014] 12 頁

⁵¹ 日本経済団体連合会[2016] 11 頁

⁵² 総務省[2017b]

こうしたデータバイアスの存在によりユーザーの全体性が把握できない場合には、既存の公的統計や他のビッグデータとの比較において、矛盾や不整合を検知する技術の研究・開発から始める必要がある。

4.3.3. データ形式・メタデータの課題

ノイズやバイアス以外の課題として、データ形式が挙げられる。通常、民間企業が保有するデータに用いられる用語や様式は公的統計のそれとは異なっているほか、同業種の企業間ですら統一されていないことも多い。従って、公的統計として活用する場合にはデータの再分類やコードの統一化などが必要となる。加えて、ビッグデータは、多くの非構造化データを含むため、これらの技術的な取り扱いも検討課題となる。

データ解析に不可欠なメタデータが僅少であるビッグデータも多い。英国が位置情報付 Twitter を利用して人口推計を試みた事例では、iPhone ユーザーの投稿だけが特定の時期を境に減少したが、一般に iPhone ユーザーが Android ユーザーに比べて平均年齢及び社会的地位が高いことを踏まえると、この時期を境に集計結果の統計上のインプリケーションが変化した可能性が指摘⁵³されている。属性情報が統計的解析を行う上でいかに重要であるか例証している。そもそも、顧客属性に関しては、本人確認書類の提示が不要な電子マネー等では、真実性の検証⁵⁴も重要となる。

SNS 情報については、フレーミングやトローリング⁵⁵のようなメッセージが拡散する現象などもある。こうしたデータから適切に統計的な結果を得るためには、ビッグデータの取り扱いを巡る「特別な経験・知見」の活用が求められよう⁵⁶。

⁵³ Sweier, Komarniczky & Clapperton[2015] 18 頁

⁵⁴ 同様の問題は SNS にも存在しており、Upadhyaya & Kynclova[2017] 18 頁では、年齢や性別を事前に記入する SNS においても、ユーザーが真実を記入しているか否かは不明であると指摘している。

⁵⁵ フレーミング (flame、炎) やトローリング (troll、釣り) は、SNS などのインターネットコミュニティ上の攻撃的な言辞のこと。

⁵⁶ Upadhyaya & Kynclova[2017] 18 頁

4.3.4. 実務上の示唆

ノイズやバイアス等を巡るデータマネジメント上の課題はデータ毎によって異なるほか、こうした課題に対処可能な技術面のノウハウも統計作成当局によって区々である。しかし、ウェブスクレイピングによるビッグデータの活用における先駆者であり、他の統計作成当局に比べて多くのデータサイエンティストを抱えるオランダでさえ、物価指数への活用に5年もの期間を要した事例⁵⁷もあり、ビッグデータの活用に向けて克服すべき実務上の課題がいかに多いかを例証している。

従って、ビッグデータ活用の検討を始める際は、「大規模に取り組むのではなく、試行的に実証実験から開始する」ことが賢明であると言える。

こうしたデータマネジメント上の課題を解決するための工夫として、外部の専門的知見を有する関係者と積極的に協働することも有益となる。例えば、データのクリーニングやデータ解析技術の研究に際しては、専門的なノウハウを有する第三者機関へのアウトソースも選択肢となる⁵⁸ほか、民間データホルダー自身が、保有するビッグデータのサンプリングや集計を施した上で統計作成当局に提出する手法⁵⁹も考え得る。さまざまな選択肢を検討しつつ、統計作成当局自身が知見とノウハウを積み上げる努力が不可欠である。

先行する諸外国の事例を踏まえると、予期せぬコスト負担が強いられることも多い。例えば、既に多くのビッグデータ活用の試行的な取組を行っている Eurostat は、大量のデータ処理に係る負担に加えて、既存の調査を継続する場合には「二重の負担」が強いられる点などを指摘しつつ、コスト削減効果について過度な期待は禁物である⁶⁰と釘を刺している。

試行的な取組として着手する場合には、実証実験を行うために必要なデータ量が多くなくても十分な場合も多く、そうしたケースでは企業にとって利用価値の高い最新のデータでなくてもよい（例えば3~7年前の5年間のデータのみ取得するなど）場合もあるため、かかる工夫を行うことでコストを抑えることも可能と思われる。

⁵⁷ Bean[2016]

⁵⁸ Struijs [2016] 20 頁

⁵⁹ Struijs [2016] 19 頁

⁶⁰ Eurostat[2017]

また、統計作成当局による検討の過程で、予期せぬ課題への直面などから対象データを変更する場合や、プロジェクトそのものを取りやめる場合にも、試行的な取り組みから始める方が、統計作成当局が対外的に負う説明責任の負担を軽減できるものと考えられる。

4.4. ビッグデータを保有する民間主体とどのように協力関係を構築するか

ビッグデータには、誰もが利用可能な一般に公開されたデータもあるが、民間主体が一般には公開していない機密性の高いビッグデータも多い状況にある。従って、データへの実効的なアクセスが課題となる。

統計作成にとっていかに魅力的なビッグデータであっても、データを取得できない場合には、試行的な実証実験にさえ取り組むことはできないため、データアクセスを巡る問題が利活用上の最大の障害要因⁶¹とみる向きもある。こうした課題に、統計作成当局はいかに向き合うべきか考えてみたい。

4.4.1. 協力の義務づけの妥当性

データアクセスを巡っては、「民間主体にデータ提供を義務づける」法制面の手当てが選択肢となる。これは、データを保有する民間主体が統計作成当局へのデータ提供に後ろ向きである⁶²ことが多く、「任意」でのデータ提供が期待し難いためである。

物価統計の作成に民間のスキャンデータを活用しているスウェーデンの事例においても、データ提供を義務化するような法制面での手当てが品質の高い指数作成に寄与すると指摘⁶³されている。

この点、データを保有する民間主体からも、任意でのデータ提供ではなく義務づけを選好する声も存在する。こうした向きは、経営戦略上の高価値資産であるビッグデータを、自社だけが任意に提出する場合には競争力の観点から不利に作用するが、競争相手にも提出を義務づけるような場合には「同じ土俵」(level playing field)が確保される⁶⁴と考えている。自社だけがデータ提供を行う場合に、競争力の観点から不利に作用するといった指摘は、統計作成当局が利用する場合であったとしても、何らかのかたちで自社の情報が外部から

⁶¹ 例えば、United Nations Statistical Commission[2015] (17頁)は、データアクセスを利活用上の最大の隘路として指摘している (While most respondents recognize the challenges related to IT, skills, legislation and methodology, the biggest challenge for most Big Data projects is the limited or restricted access to potential datasets.)。

⁶² Eurostat[2017] 22頁

⁶³ Sammar[2013] 3頁

⁶⁴ 英国王立統計学会の Shah 事務局長による英国議会でのヒアリングでの発言。詳細は、House of Commons[2016]を参照のこと。

識別され得るリスクを警戒しているものと考えられる。

民間主体が任意でのデータ提供に躊躇する背景について改めて考えてみると、

- ① ビッグデータは経営戦略上の高価値資産であり提供インセンティブが僅少であること
- ② 公的当局への提供を巡る法的不確実性があること

に集約することができよう。従って、法制面の手当ては、データの「任意」提供を巡る課題を克服できる可能性がある。

しかし、統計調査の多くが、調査客体たる家計や企業等の民間主体との信頼関係を基盤として実施されている事実を踏まえると、ビッグデータという新たなデータ源について幅広く提供を義務づける必要があるかは、慎重に考える必要がある⁶⁵。

ビッグデータの場合にはデータマネジメントを巡って大きな課題が存在しており、ノウハウを有する第三者の知見やノウハウの活用が非常に有効である。データ提供を受けた後も幅広い官民協働が不可欠であることを見据えると、データ提供に限って義務づけを行うことで、却ってその後の協力を阻害する懸念はないだろうか。これまで築いてきた統計作成当局と民間主体との信頼関係が変化するリスクもある。

さらには、民間主体自身によるビッグデータ活用が活発化する中で、データ提供の義務づけが統計作成当局から将来的に提供される成果物への過度な期待につながり、先進的な民間統計の作成や、ビジネスを目的としたデータ活用の発展を妨げる不安も残る。協力の義務づけの妥当性を巡っては、こうしたリスクを十分踏まえた上で丁寧な議論がなされることが期待される。

4.4.2. 対価を払ったデータ提供

民間主体から「任意」でデータ提供を受ける場合、「金銭での対価を払う方法」が有力な選択肢の一つとなる。

⁶⁵ 統計作成のために家計や企業にデータ提供を求める法律がある中でも「法的根拠に過度に頼るべきでない」との方針を掲げたルクセンブルクの事例などが存在する。詳細は、Hury & Lamboray[2013] 3頁を参照のこと。

ビッグデータの統計的利活用を推進するために、総務省が幅広い民間企業等を対象に実施したヒアリング（図表 7）でも、「企業におけるデータ収集コストの一部は、ユーザーに転嫁している。無償提供を前提とする場合には、疑問が残る」との意見が聞かれており、データ提供に見合う費用を統計作成当局が負担する場合には、民間主体の理解が得られやすいものと思われる。我が国では既に、有償購入した POS データを一部活用するかたちで物価指数を作成している（前掲図表 1）。

なお、データ購入以外にも、データ提供者に対する税制優遇措置⁶⁶なども一つのアイデアとして提案されており、民間主体への金銭インセンティブの方法はさまざま考えられよう。

図表 7：総務省ヒアリング結果（抜粋）

<金銭インセンティブ関連>

- ◆ 企業におけるデータ収集コストの一部は、ユーザーに転嫁している。無償提供を前提とする場合には、疑問が残る。

<統計作成当局のノウハウ関連>

- ◆ データの性質次第で提供できる民間データも多いはず。例えば、データ量が少ない取組では、他社とデータを出し合う動きもみられ、国が橋渡しを行うことも可能ではないか。
- ◆ 売上データの分類方法・カバー範囲など、企業間で異なるデータを比較・統合可能な形式にデータクレンジングを行う必要。統計行政が持つノウハウ活用や分類の標準化などに期待。

<企業イメージ関連>

- ◆ 企業が無償でデータを提供する場合には、それが企業や提供サービスに対する認知度の向上となるか、もしくは研究機関等とのつながりが持てるか、等が理由になる。何らかのインセンティブ設計が必要。
- ◆ 利活用の目的について、公益性の高いものであれば、CSR の観点から参画する企業も多いだろう。
- ◆ 各府省から内々に依頼されるより、正式なプロジェクトとして要請された方が社内決裁もスムーズに運ぶだろう。
- ◆ 協議会に参画する思惑は企業によって区々ながら、企業イメージや認知度を上げる効果は等しく期待されよう。

<自社サービス関連>

- ◆ 同業の競合相手にデータを出すことはできないが、異業種間ではデータを見せ合う可能性はより期待される。協議の場を設ける場合には、参画企業の顔ぶれも重要な要素。

（注）筆者が設けた見出しに再分類している。

（資料）総務省[2017b]をもとに作成。

⁶⁶ 内閣府[2015]、経済同友会[2016]

4.4.3. 統計作成当局のノウハウや成果還元

前掲の総務省ヒアリングでは、金銭での対価のほかに、研究機関等とのつながりや他社とのデータ共有における国の橋渡し機能、企業保有データのクレンジングや分類の標準化などを期待する声が聞かれており、「統計行政が持つノウハウの提供」がデータ提供のインセンティブとして機能することが窺われる。

物価統計にスキャンデータを部分活用しているルクセンブルクでは、データ提供と引き替えに「データ提供企業向けに個社データとマクロ動向を比較分析したレポート」を作成して還元する取り組み⁶⁷がみられている。これに関連して、データ提供者向けのワークショップを開催するアイデア⁶⁸も提案されている。

また、ベルギー統計局と Eurostat が、携帯キャリア大手の Proximus 社と 2015 年に開始したジョイント・プロジェクトでは、「三者相互のデータアクセス」を認めることを通じて、各々が保有するデータだけでは解析できない現象を把握するなど、WIN-WIN なパートナーシップを構築した事例⁶⁹として注目されている。

このように、統計作成当局のノウハウや成果還元は、民間主体が統計作成当局にデータを提供する際の大きなインセンティブとなり得る。特に、特定の民間保険会社が保有する加入者の健康状態記録から、全人口のトレンドを捕捉可能か否か、オンライン物価が一般物価全体を捉え得るか否かなど、統計作成当局とデータを保有する民間主体が「共通の悩み」を共有できるような場合には、官民協働によるデータセットの整合性の調整は共通のインセンティブ⁷⁰となる。

4.4.4. 民間企業のイメージ向上

データを保有する民間主体が、「自らの社会的なイメージの向上」を図る目的で、統計作成当局からのデータ提供の要請に応じる可能性がある。この点、前掲の総務省ヒアリング（図表 7）でも、「公益性の高いものであれば、CSR の観点から参加する企業も多いだろう」との声が聞かれている。

⁶⁷ Hury & Lamboray[2013] 3 頁

⁶⁸ Sammar[2013] 3 頁

⁶⁹ Klein[2017] 12-13 頁

⁷⁰ Landefeld[2014] 13-14 頁

こうした企業の参画を一段と促す工夫として、ビッグデータの提供企業に「公益性の高い企業」として認定するアイデア⁷¹なども提案されており一考に値する。

そもそも、近年では、民間企業や業界団体などが自ら CSR の一環として、保有するビッグデータの活用を通じて広く公益に資するような統計的分析を実施する先進事例も少なくない。例えば、ある米大手商業銀行が、同行顧客の口座記録をもとに行った貸金ボラティリティの原因分析⁷²があるほか、我が国では、一橋大学と新日本スーパーマーケット協会、インテージ社が開発した SRI 一橋大学消費者購買指数・単価指数⁷³など、「社会貢献のために無償で行われる」取り組みが存在している。

こうした社会的意義の大きい民間の取組に、統計作成当局がどのように向き合い、協働していくかも重要な論点の一つである。統計ニーズに応じたさまざまな民間統計が生み出されることで、公的統計として活用可能性のあるデータ量が拡大するメリットが考えられる。他方で、フェイクニュースがソーシャルメディアに蔓延るように、フェイク統計が品質の高い統計を駆逐するリスク⁷⁴の顕在化を防ぐような戦略的な取組が求められる。

こうした観点からは、民間主体が自主的に統計作成に取り組む際に、既存の公的統計との品質等の相違についてのメタデータの提供に統計作成当局として協力する方法も一案となる。さらには、統計作成当局が、統計の作成主体としてではなく、民間主体が作成した統計の「クオリティチェック」(accreditation / certification) の役割を果たすアイデア⁷⁵も提起されており、今後、統計作成当局全体として考えるべき大きなテーマの一つであろう⁷⁶。

⁷¹ 内閣府[2015] 4 頁

⁷² Farrell & Greig [2016]

⁷³ スーパーマーケット、コンビニエンスストア、ドラッグストア等の業界から全国約 4 千店舗の業務データを収集し分析。詳細は、SRI 一橋大学消費者購買指数・単価指数ホームページ (<http://www.srihito.com/index.html>) を参照。

⁷⁴ Cœuré[2017]

⁷⁵ MacFeely [2016] 793 頁, Hammer *et al.* [2017] 19 頁

⁷⁶ Struijs [2016] (23-24 頁) では、民間企業自らが積極的にデータ活用を行う中で、公的統計に対する社会の依存度が低下していると指摘した上で、統計作成当局が統計作成の任を離れて、第三者が作成した統計の品質を評価する役割を果たすべきかとの論点を提示。ほか、Florescu *et al.*[2014] (10 頁) では、情報過多の社会 (jungle of information) にあって質の担保が必要との時代認識の下、統計作成当局は、統計作成の独占的地位を有している訳でもデータ社会の警察でもないが、長い歴史の中で統計品質に係るモラルを有してきた

4.4.5. 民間企業のサービスの向上

民間ビジネスにおけるビッグデータの統計的利活用の状況を俯瞰すると、主に内外の金融市場参加者向けに民間統計を開発の上で有償販売する取組も散見されている。こうした商品については、統計作成当局との連携や、政策当局による積極的な利活用（例えば、財政政策、金融政策の前提となる景気判断の材料の一つとして活用する等）が宣伝効果に繋がる可能性がある。

この点、政府が提供する RESAS が「ビッグデータの見本市」のような役割を果たすとの指摘⁷⁷もみられており、統計作成当局が民間データの受け皿を作ることを通じ「商品の認知度向上」を企図した民間主体の無償提供を促しつつ、統計作成当局が利用できるデータの拡大を図る方法も考え得る。

ただし、先に述べたとおり、統計作成当局が積極的に民間統計を活用する場合には、統計としてのクオリティチェックが不可欠である点には注意が必要である。民間統計がビジネスの一環として提供される場合でも、先進的な民間統計について「政府認定」することで利活用を促すアイデアも選択肢の一つとして検討できる⁷⁸。

国土地理院が作成・提供する地形図の事例ではあるが、民間地図のベースとして利用されている地形図の作成において、登山アプリ業者が保有する登山者の移動経路情報を精度向上に役立てる取組⁷⁹が進捗中であり、「民間企業の商品ニーズと当局の施策ニーズを上手くマッチングさせる」ことで民間主体によるデータ提供が進む可能性を例証している。民間主体のニーズを踏まえた多様な連携のあり方が考えられよう。

4.4.6. データ取得の継続性を巡る課題

統計作成当局の立場からみると、民間データ及び民間統計の精度検証のためには良質なデータを相応に確保することが必要となる。

と評価し、データ作成者に対する承認制度に触れている。

⁷⁷ 経済同友会[2016] 41 頁

⁷⁸ 柳川・渡辺[2017]では、統計作成サービスの民間開放について論じており、理論的なバリエーションの一つとして、特定の民間提供の統計を政府統計として認定するケースを想定。その際は、一定のルールと監督体制を整える必要があるため、早い段階から進めるべきと結論づけている。

⁷⁹ 国土交通省国土地理院[2018]

例えば、トレンド分析を行うためには、過去に遡ってデータを収集して時系列指標を作成する必要があるほか、当該ビッグデータを統計作成に利用するためには、「将来に亘っても安定的にデータを確保する」ことが課題となる。公的統計の長所の一つとして、時系列データの充実が指摘⁸⁰されるが、長期的なデータ確保が達成されない場合には、公的統計としての強みが大きく制約されることとなる。

データの継続性に係る課題を巡っては、主に以下の状況が懸念される。

- ① データを保有する民間主体が、過去のデータを蓄積していない場合
- ② ビジネス環境の変化によりデータ量が増減する場合
- ③ 民間主体がデータ提供を取り止める場合

①については、おもに一般に公開された情報を取得する場合に想定される。例えば、宿泊施設の予約や求人サイトでは宿泊日や募集期間等を過ぎたデータがウェブ上から消去されてしまう問題が生じる。

こうした問題に対処するためには、統計作成当局がデータ蓄積のための基盤⁸¹を講じるなどの対応が求められよう。

②については、データを保有する民間主体の業界における市場シェアの変動や、データ取得に関する環境変化（パーソナルデータに係るプライバシーポリシーの変更など）が想定される。例えば、位置情報付きの Twitter 情報を活用して人口変化を推計した英国の試みでは、iPhone のオペレーティングシステムの変更（プライバシーポリシー等）後に利用可能データが 25% 程度減少した事例があり、データをコンスタントに確保できないリスクを例証している。

こうした問題へは、得られたデータから把握できる動向のみを額面通りに受け止めるのではなく、データ変化の背景にある技術革新や社会情勢の変化まで視点を広げて分析⁸²することで対応を図ることが重要である。

⁸⁰ Eurostat[2017] 24 頁

⁸¹ 一藤・曾根原[2016]

⁸² Sweier, Komarniczky & Clapperton[2015] 6 頁

③については、特定の民間主体と協力関係を構築した場合でも、当該民間主体が突然データ提供を停止するリスク⁸³が存在する。外部のデータ源へ依存度を高めることは、統計作成当局による統計作成を巡る自主的なコントロールが低下する欠点がある。

こうしたリスクを軽減する観点からは、取得する民間データのバイアスが上手く補正可能な場合でも、敢えて複数社からデータ提供を受けるなどの工夫が考えられる。このほか、データ提供を受けるに際して民間主体と統計作成当局の間で連携協定を締結することで、一定期間のデータ提供を確保する方法も考えられる⁸⁴。

⁸³ Abdulkadri, Evans & Ash[2016] 26 頁

⁸⁴ 例えば、スーパー小売各社がスキャンデータの任意提供を取りやめる場合には、統計作成当局に 6～12 か月前に申し出ることを求める通知期間（notice period）を定めたスウェーデンの協定の事例が存在する。詳細は、Sammar, Norberg & Tongur[2013] 3 頁を参照。

4.5. 対外コミュニケーションのあり方はケース・バイ・ケース

ビッグデータを保有する民間主体の考え方はさまざまであるため、望ましい協力体制は各主体とのコミュニケーションを踏まえてケース・バイ・ケースで判断する必要がある。また、データを保有する民間主体以外の関係者へのコミュニケーションも重要である。ここでは、いくつかの先行事例を踏まえつつ、統計作成当局がコミュニケーションを実施する上で参考となる工夫と効果について具体例を交えながら説明する。

4.5.1. データ保有主体へのコミュニケーション

データの提供者たる民間主体と統計作成当局の信頼関係がない場合には、任意で安定的にデータを確保することが困難となる。従って、統計作成に必要なデータ量を確保する以上に、コミュニケーションによる信頼関係の醸成が先決⁸⁵と言える。

企業ホームページ情報など一般に公開されたデータを活用して統計作成を試みる場合にも、データを大量に収集する場合には事前のコミュニケーションが有効となる。オランダの事例では、企業ウェブサイトのスクレイピングを実施する前にサイトオーナーに接触した結果、スクレイピングを実施せずとも（当該企業のバックオフィスから）直接データ提供を受けたケース⁸⁶がある。反対に、英国の事例では Twitter 社が提供する API を通じてデータを（一般に公開されたものとして）取得していたが、同社のコンプライアンス⁸⁷上の疑問が指摘されたことから、プロジェクトの途中からデータを有償購入に切り替えた事例⁸⁸がある。

いずれの事例も、本格的な取組を開始する前にデータを保有する民間主体と積極的な対話を行うことの重要性を例証している。

⁸⁵ ルクセンブルクの事例では、少量のデータ提供から始めることで信頼関係を構築し、かかる基盤をもとにして更に必要なデータ提供を要請する方が、信頼関係がない状況で一度に多くのデータを求めるよりも全体として近道であると指摘されている。詳細は、Hury & Lamboray[2013] 3 頁を参照のこと。

⁸⁶ Bosch & Windmeijer[2014] (8 頁)

⁸⁷ The Twitter developer Rules: Section 1C “If your application will need more than 1 million user tokens, you must contact us about your Twitter API access, as you may be subject to additional terms.”

⁸⁸ Swier, Komarniczky & Clapperton[2015] 13-14 頁

4.5.2. データ保有主体以外への対外コミュニケーション

民間主体からデータ提供を受ける場合には、「データ保有主体以外」へも透明性の高いコミュニケーションを行うことが重要である。

前掲の総務省ヒアリングによると、データ提供に関して「各府省から内々に依頼されるより、正式なプロジェクトとして要請された方が社内決済もスムーズに運ぶだろう」との声が聞かれている。当該民間主体以外への公表を見据えたコミュニケーションを行うことで、データ保有企業の参画を促すことができるだろう。

また、民間主体がデータ提供を躊躇する背景の一つとして、データ提供を巡る法的不確実性の存在があることは既に述べたが、この背景には、データを保有する企業が自らの顧客や投資家などに対外的な説得材料を必要としていることも一因となっている。従って、統計作成当局が、こうした企業側の懸念を払拭するような、対外的な広報を行う工夫が求められる。

この点は、データ提供の要請を行う統計作成当局にとっても重要である。とりわけ、特定企業から任意でデータの無償提供を受ける場合には、対外広報をしっかりと行うことを通じて公正な統計行政に対する不要の疑念を惹起しないよう努めるべきほか、データ提供企業以外にも取組への将来的な参加の途を確保するなどの配慮が求められよう。統計作成当局が行う統計調査は、あらゆる民間主体との信頼関係が基盤となっているため、これまで築き上げた信頼を失わないようなコミュニケーション上の努力が必要である。

もっとも、データ保有主体の側が、対外コミュニケーションを望まない場合が存在する点には注意すべき必要がある。一例として、カナダがスマートメーターの活用を検討した際には、配電事業者の名称は非公開とされ、統計局による報告書においても「A社」や「B社」などと呼称された事例⁸⁹などがある。このような場合には、データ保有主体と統計作成当局との間で、対外広報を巡るスタンスを摺り合わせる必要がある。

⁸⁹ Ma[2014]

4.5.3. データコミュニティの深化を促す対外コミュニケーション

統計作成当局による試行的な取組においては、最終的な成果物としての公的統計だけでなく、各取組の進捗状況を適宜公表することが望ましい⁹⁰。検討プロセスを透明化することのメリットについて達観すれば、以下の4点が期待される。

- ① 各種課題やテーマごとに精通した第三者からの知見・ノウハウの提供が期待できること
- ② 統計ユーザーからのフィードバックが期待され、さまざまな統計ニーズを踏まえて検討を実施できること
- ③ 優良事例のみならず失敗事例を蓄積することで、将来の取組においてより効率的な検討を実施できること
- ④ これまで公的統計のコミュニティの外にあったような企業や専門家などの関心を高め、幅広い協力を引き出すこと

統計作成当局は、これまでも統計学に知見を有する学術研究者とも協力しながら公的統計の整備に努めてきた。今後、ビッグデータを公的統計のデータ源として活用するためには、これまで以上に幅広いデータコミュニティを対象に連携を深化させるよう努める必要がある。

⁹⁰ 英国国家統計局が取り組んだ4つのパイロット・プロジェクトでは、半期に一度進捗レポートを公表している。当該レポートでは、検討中に直面したさまざまな課題や制約についても積極的に公開している。

4.6. 個人情報や企業情報の保護

ここからは、ビッグデータの利活用目的やデータを問わず重要な分野横断的な論点について考えてみたい。

第一は、「個人情報や企業情報の保護」を巡る論点である。民間主体のデータ提供に伴う不安を払拭するためには、提供を受けたデータをしっかりと保護する措置を講じる必要がある。統計作成当局にとっても、統計調査が調査客体たる家計や企業等との信頼関係を基盤として行われている事実を踏まえると、民間ビッグデータについても統計調査における調査票情報の管理と同様に秘密の保護には万全を期す必要がある。

かかる観点からは、公的統計に民間主体のデータを活用した際のデータの所有権（data ownership and copyright⁹¹）を巡る取り決めも論点となるほか、データ授受に関するサーバーセキュリティリスクへの対応なども実務上の検討課題となる。

パーソナルデータ等を保有する企業にとっては、顧客への丁寧な説明も求められるため、データ提供の透明性を確保することが求められる。民間主体が行うビッグデータ活用の事例の中には、個人識別情報が含まれないデータであっても所謂炎上リスクが顕在化した事例⁹²もあり、統計作成当局がアクセス可能なデータ範囲を事前に取り決めておくなどの工夫が求められる。一例として、英国国家統計局が実施したパイロット・プロジェクトでは、取り扱うデータを（i）一般に公開されたデータ、（ii）匿名化済のデータ、（iii）集計済のデータに限定⁹³しており、慎重な態度で検討を開始している。

データのアクセスに法的制約がない場合でも、「エシカルな配慮」を怠る場合には、民間主体及び統計作成当局の双方のレピュテーションを低下させる懸念があり注意する必要がある。特に、我が国におけるパーソナルデータに係る国民の不安は、他国対比で大きいとの調査もあるため⁹⁴、国民感情を踏まえた十分な対応を整える必要がある。

⁹¹ Struijs [2016] 20 頁

⁹² Renew London 社がロンドン市に無償提供したゴミ箱から、Wi-Fi 機能を有するスマートフォンの MAC アドレスを収集していた事例（BBC News: <http://www.bbc.com/news/technology-23665490>）などがある。

⁹³ Naylor[2016]

⁹⁴ 総務省[2017b]

エシカルな観点からは、ウェブスクレイピング等を実施する際には、所謂ネチケット（netiquette）への配慮も重要であろう。オランダ統計局は、スクレイピングを実施する際には（i）対象企業が統計局ロボットによるアクセスであると確認可能な方法を採用すること、（ii）企業サイトへの負荷を軽減するためサーバアクセスの間隔を設定すること、（iii）サーバへの負荷が高まる時間帯を回避すること、といった方針⁹⁵を纏めており、サイトオーナーの立場に立って取組を実施することが重要である点を示唆している。

4.7. データ活用のための技術・人材

分野横断的な課題の2つ目として、利活用推進を支える「技術と人材の確保」が重要となる。

これまでも触れたように、大量のノイズとバイアスを有するビッグデータから、質の高い統計を作成するために必要な情報を抽出するためには、データマイニングや機械学習、AI技術など最新のデータ解析技術の研究・開発が必要となる。

そのためには、試行的な実証実験に取り組みながら統計作成当局がビッグデータ利用に必要な知見とノウハウを積み上げる努力が必要であることは既に述べたとおりである。諸外国では、更に踏み込んで、ビッグデータの解析に特化した施設を新たに設ける動きも見られている。例えば、英国国家統計局は、従前のコンピューターでは処理できない大規模なデータを取り扱うために2013年に「イノベーション・ラボ」を設立し、ここを職員の「砂場」（sand pit）として先進的な技術を試みつつ知見の蓄積を図っている⁹⁶。

統計作成当局の人材面での課題を巡っては、ビッグデータの取り扱いに係る統計的専門技術や能力を有するデータサイエンティストの育成が不可欠であるほか、これまで述べてきた各種課題に適切に対応するためには、ITや法律の専門家を含めた様々なバックグラウンドを持ったスタッフを確保する必要がある。従来のサンプリング理論をアプリオリとしない態度やマインドセットが重要との指摘⁹⁷もあり、これらを支える当局サイドの組織文化を変えていくことも

⁹⁵ Bosch & Windmeijer[2014], Hoogteijling[2016]

⁹⁶ Bean[2016] 168頁

⁹⁷ Struijs [2016] 22頁

大事な論点となる。

これまでの諸外国の事例では、3～4名のコアとなる人材でプロジェクトを立ち上げるケースが多い⁹⁸が、英国国家統計局ビッグデータチームの11名、イングランド銀行データサイエンス・ラボの12名、オランダ統計局におけるビッグデータ中核人材の5名のように、各国によって構成人員は区々である。また、オランダ統計局には中核スタッフ以外にもデータサイエンスに通じたスタッフが80名も在籍⁹⁹しており、統計作成に携わる職員全体のリテラシーの底上げ努力も忘れてはならない。

この点、Eurostatが統計職員向けに実施する統計研修（ESTP）では2016年以降、ビッグデータ関連のプログラムを充実させている。講義内容の詳細は不明ながら、機械学習やウェブスクレイピングなど特定の高度技術に特化した講義などが、2～4日間程度の集中プログラムとして年6回程度提供されている（図表8）。

図表8：ESTPにおける研修（2017年実績）

研修テーマ	期間
Introduction to Big Data and its tool	2.5日間
Machine Learning Econometrics	3.0日間
Hands-on immersion on Big Data tools	3.0日間
Big Data Sources – Web, Social Media and Text Analytics	4.0日間
Automated collection of online prices: sources, tools and methodological aspects	4.0日間
Advanced Big Data Sources – Mobile Phone and other sensors	4.0日間

（資料）European Statistical Training Program 2017をもとに作成。

我が国では、国家公務員及び地方公務員に対する統計研修業務を行っている統計研究研修所が、2017年度に初めてビッグデータに係る研修プログラム¹⁰⁰を実施したほか、2018年度にはより実践的な演習を展望した新規講義を開講する予定となっている。欧州のように、ビッグデータ活用に係る個別具体的な取組が進捗すれば、各種データや個別のプロジェクトに応じた知見や教訓などの共有を通じて、内容面での高度化も期待されよう。

⁹⁸ Hammer *et al.* [2017] 23 頁

⁹⁹ Bean[2016] 168-170 頁

¹⁰⁰ 総務省[2018b]によれば、2017年度に開講した「特別コース・ビッグデータ入門」では、40名の募集人員に対して89名の応募があったため、急遽2回に分けて開催した模様。

4.8. コーディネーション

分野横断的な課題を巡る最後の論点として、さまざまなビッグデータ活用を巡るプロジェクト間の「コーディネート」の重要性を指摘しておく。

ビッグデータ活用に向けた試行的な取組を効率的に実施するためには、たとえビッグデータの種類や参加する民間主体が異なっていたとしても、他のプロジェクトで得られた知見や技術、教訓等の成果を活用することが重要である。特に、我が国は行政ニーズに的確且つ迅速に対応するため、分散型統計機構の下で各府省が所管行政に係る統計を作成する機会が多い。この点、検討テーマやデータ、各種リソースが重複しないよう、「各府省による取組を束ねる鳥瞰的な視点」が不可欠である¹⁰¹。第3節で紹介した第Ⅲ期基本計画では、各府省における効率的な活用を推進するため、統計委員会を中心に情報共有を図る施策等が盛り込まれている（前掲図表3）。こうした具体的な施策の着実な実行が期待される。

長期的な視点として、「国際的なコーディネート」のあり方も検討課題となろう。現状、統計作成当局が行う調査統計については、統計の質や調査項目の定義、分類などについて国際標準が定められる場合が多く、従って、作成された統計の国際比較が容易である点が、公的統計の強みの一つとなっているが、各国独自にビッグデータ活用に取り組む場合、作成された統計の国際比較は容易ではない。国内における利活用上の課題だけでなく、国際的に検討すべき論点についても今後取り組む必要がある。

¹⁰¹ 総務省統計委員会[2017a]（第77回基本計画部会・第9回横断的課題検討部会）によると、西村統計委員会委員長は「最初の段階ではこういう形で、それぞれの省が、自分のイニシアチブでいろいろなことをやっていただくというのは良いのですが、どこかでコーディネーションをしなければいけないという感じがします。（中略）これについては統計委員会としても少し考えなきゃいけない」と発言している。

5. 結びにかえて

本稿では、公的統計へのビッグデータ活用のメリットと先行事例を簡単に整理した後、我が国のビッグデータ活用の本格化に向けた近年の統計改革における検討状況を紹介した。我が国においても、ビッグデータの統計的利活用の推進に期待が高まっており、総務省においてデータの相互利活用を促進するための産官学連携による協議の場が設けられるなど、具体的な取組を後押しするような政府レベルでの施策が実行段階に移りつつある。

しかしながら、我が国におけるビッグデータ活用に向けた検討は緒に就いたばかりであり、実務上の検討プロセスも定まっていない状況にある。そこで本稿では、将来的な活発な取組を促す観点から望ましいと考えられる雛形を一つ提示し、諸外国の先行事例に学びながら、ビッグデータを保有する民間主体との協力関係の構築のあり方を中心に実務上の含意を探った。

ビッグデータの活用に向けた検討は、これまでの公的統計の作成プロセスとは大きく異なるため、まずは、目的を明確化した上で適切なデータを特定し、試行的な実証実験を繰り返すほかない。「ビッグデータは特効薬ではない¹⁰²」との指摘もあるように、データマネジメントやデータアクセスの課題など目的達成のためにクリアすべき検討課題も多い。これらを支える統計的技術や人材、さまざまな取組を束ねるコーディネーション機能も不可欠である。

本稿が示した検討プロセス及び課題解決のための工夫は、あくまで統計作成当局がビッグデータ活用の検討を開始する際の手掛かりに過ぎない。今後は、さまざまな試行的な取組を通じて、より効率的且つ効果的にビッグデータの統計的利活用を実現できるようなベストプラクティスを見出していくことが期待される。

以 上

¹⁰² Bean[2016] 172 頁

補論：ビッグデータ活用に向けた国際的な取組の概観

統計作成当局によるビッグデータ活用の検討は、オランダ統計局が1992年には既にスーパーマーケットのスキャンデータの取得に着手する¹⁰³など、一部の統計局において散発的にみられていたが、経済社会におけるデータ量が爆発的に増加するもとの、近年では国際機関などにおいても戦略的かつ体系的な取組が推進されつつある。

ここでは、国際連合、欧州統計システム、国際通貨基金、各国中央銀行を取り上げて、ビッグデータ活用に係る取組概要を紹介する。

1. 国際連合

国連統計委員会は、2014年3月にビッグデータ活用に向けたグローバル・ワーキンググループを設置した。

計8つのタスクチームを設けており、衛星画像やスキャンデータ、携帯電話情報、ソーシャルメディア情報などの個別のデータに係るタスクを設けつつ、データアクセスやSDGs（持続可能な開発目標）への応用可能性、データスキル、国際ネットワークの構築などの分野横断的なテーマも設定している。メンバーシップは、2008年3月時点で26の国連加盟国と世界銀行やEurostatなど14の国際機関となっている¹⁰⁴。

図表：国連におけるタスクチーム

1.	データアクセスやパートナーシップ締結の促進
2.	SDGs指標への活用可能性
3.	衛星画像・地理空間情報
4.	携帯電話情報
5.	スキャンデータ
6.	ソーシャルメディア情報
7.	ビッグデータを扱うためのスキルやトレーニング
8.	人材や知見共有の国際ネットワークの構築

(資料) 国連ホームページをもとに作成。

¹⁰³ Bean[2016] 170頁

¹⁰⁴ United Nations Statistical Commission[2018]によると、参加国は、オーストラリア、バングラディッシュ、ブラジル、カメルーン、カナダ、中国、コロンビア、デンマーク、エジプト、ドイツ、インドネシア、アイルランド、イタリア、メキシコ、モロッコ、オランダ、オマーン、パキスタン、フィリピン、韓国、サウジアラビア、スイス、アラブ首長国連邦、英国、タンザニア、米国。現時点で日本は未参加。

2. 欧州統計システム

Eurostat や各国統計局からなる欧州統計システム (ESS) では、2013 年に、「ビッグデータと公的統計のスヘフェニンゲン覚書」を締結し、これを受けて 2014 年には Eurostat を中心に設けられたタスクフォースチームにより、覚書を実行に移すために必要なアクションプランとロードマップを策定した。

2016 年 3 月には“ESSnet Big Data” と呼称されるプロジェクトを立ち上げ、2018 年 5 月を目途として、個別のデータに着目した 7 つの具体的なタスクと、それらを支える全体にかかる 3 つのタスクからなる 10 のワークパッケージ (WP) に取り組んでいる。

図表：ESS が実施するワークパッケージ

WP 0	コーディネーション	WP 5	携帯電話データ
WP 1	求人情報のスクレイピング	WP 6	公的統計の早期予測
WP 2	企業情報のスクレイピング	WP 7	人口・観光・農業の統計分野
WP 3	スマートメーター	WP 8	統計技術・IT 技術
WP 4	船舶の位置情報 (AIS)	WP 9	公表

(資料) 欧州委員会ホームページをもとに作成。

3. 国際通貨基金

国際通貨基金 (IMF) は、2015 年 11 月に「ビッグデータ・イノベーション・チャレンジ」を立ち上げており、6 つの先駆的なアイデアの実証実験に取り組んでいる。

ホテルの宿泊料やタクシー代金などのビッグデータを活用し、ビーチホリデー観光に係る価格指標¹⁰⁵を作成する取組や、SWIFT¹⁰⁶データを活用した国際的な資金フローの把握など、国際比較やグローバルな動向把握を可能とする統計作成への取組が中心である。

¹⁰⁵ Week @ the Beach Index は、英 The Economist 誌によって公表されるビッグマック・インデックスから着想を得て発明された、ビーチホリデー観光に係る価格指標のこと。3 つ星ホテルの宿泊料やタクシー代金、飲食費等からなるバスケット価格を国際比較する。IMF では、Travelocity 社や Trivago 社のビッグデータなどを活用し、同指数の改善に取り組んでいる模様 (Hammer *et al.* [2017])。

¹⁰⁶ SWIFT (Society for Worldwide Interbank Financial Telecommunication) は、銀行間の国際金融取引に係る事務処理の機械化、合理化および自動処理化を推進するため、参加銀行間の国際金融取引に関するメッセージをコンピューターと通信回線を利用して伝送するネットワークシステムのこと。

図表：IMFによるイノベーションチャレンジ

1.	SWIFTデータを活用した国際資金フローの把握
2.	早期警戒システムの構築のためのセンチメント把握
3.	Googleトレンドを活用したGDPのナウ・キャストイング
4.	Week @ the Beach Indexの拡大
5.	政策分析のための政府支出データの蓄積
6.	税務行政に資するデータ分析の応用

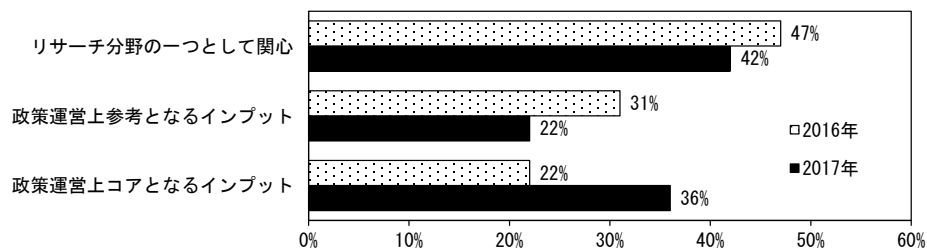
(資料) Hammer *et al.*[2017]をもとに作成。

4. 各国中央銀行

物価安定や金融システム維持の役割を担い、統計作成の任も負う中央銀行においても、ビッグデータ活用を通じた景気動向のタイムリーな把握に取り組んでいる。中央銀行を対象としたサーベイ¹⁰⁷では、近年、各国中銀によるビッグデータに対するスタンスが変化しており、研究・調査活動における関心を超えて、政策運営にとって重要なデータ源として取り扱われるようになりつつある。

また、中央銀行の統計専門家によるアービング・フィッシャー委員会¹⁰⁸においても、中央銀行とビッグデータをテーマとする会合が開催されている¹⁰⁹。

図表：各国中銀によるデータ源としてのビッグデータの捉え方



(資料) Glass[2017]、Central Banking[2016]をもとに作成。

¹⁰⁷ Glass [2017]、Central Banking [2016]

¹⁰⁸ 中央銀行の統計に関する諸問題を討議すべく、中央銀行の統計専門家により1997年に発足し、2006年よりBIS (Bank for International Settlements、国際決済銀行) が事務局を担っている委員会。日本からは日本銀行が参画。

¹⁰⁹ 2017年に行われた会合では、アルメニア中銀によるウェブ情報を活用した物価・不動産・雇用動向の把握や、スウェーデン中銀によるネット価格情報を活用したインフレ予測、オランダ領アルバ中銀によるネット検索情報を活用した観光収支の計測、ドイツ連銀によるネット検索情報を活用した預金者センチメントの分析事例などが共有された。詳細は、Irving Fisher Committee[2017]を参照。

参考文献

會田雅人 [2015] 「ビッグデータ時代と公的統計」、『統計』2015年9月号

一藤裕・曾根原登 [2016] 「ソーシャル・ビッグデータ駆動の観光政策決定支援システム」、『システム／制御／情報』Vol.60、No.4

閣議決定 [2018] 『公的統計の整備に関する基本的な計画』

北村行伸 [2018] 「統計リテラシーとは何か」、『日本統計学会会報』No.175

国土交通省国土地理院 [2018] 「ビッグデータで登山道を修正した地形図をはじめて公開」、2018年3月7日付報道発表資料、国土交通省国土地理院基本図情報部

高田悠矢・別所英実・五十嵐盛仁 [2018] 「労働市場の民間ビッグデータ：経済統計としての活用可能性」、総務省統計委員会担当室ワーキングペーパー、2018-WP02

経済財政諮問会議 [2016] 『統計改革の基本方針』

経済産業省 [2017a] 『平成28年度IoT推進のための新産業モデル創出基盤整備事業（ビッグデータを活用した新指標開発事業）報告書』、経済産業省大臣官房調査統計グループ調査分析支援室委託調査、PwC あらた有限責任監査法人

—— [2017b] 「ビッグデータ・AIを活用した新しい経済指標への意見を募集します～既存の統計よりも優れた速報性・詳細性が期待されます～」、2017年7月19日付ニュースリリース、経済産業省大臣官房調査統計グループ調査分析室

経済同友会 [2016] 「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－」、経済同友会

総務省 [2017a] 「インターネット通信販売価格の把握に関する平成28年度実施報告」、物価指数研究会（第8回）書類番号1、総務省統計局物価統計室

—— [2017b] 『平成29年版 情報通信白書』、総務省

—— [2017c] 「モバイルビッグデータを活用したテレワーク・デイの効果検証」、2017年10月13日付報道資料、総務省情報流通行政局情報流通振興課情報流通高度化推進室

- [2017d]「消費動向指数研究協議会の設立」、2017年7月28日付報道資料、統計局統計調査部消費統計課
- [2018a]「ウェブスクレイピングによる宿泊料データの取集と分析用データの抽出」、物価指数研究会（第10回）書類番号2、総務省統計局物価統計室
- [2018b]『総務省広報誌 Vol.209』2018年5月号、総務省
- [2018c]「ビッグデータ等の利活用推進に関する産官学連携のための連携会議の開催について」、2018年5月22日、総務省
- 総務省統計委員会 [2017a]「第77回基本計画部会・第9回横断的課題検討部会（合同部会）議事録」、総務省
- [2017b]「基本計画部会・第9回共通基盤ワーキンググループ会合資料2-1」、総務省
- 統計改革推進会議 [2017]『統計改革推進会議最終とりまとめ』
- 内閣府 [2014]「消費税率引上げ後の消費動向等について」、2014年4月～11月、内閣府政策統括官（経済財政分析担当）付参事官（総括担当）
- [2015]「公的統計におけるビッグ・データの活用に関する調査研究」、平成26年度内閣府大臣官房統計委員会担当室請負調査、株式会社NTTデータ経済研究所
- [2018]「テキストデータを用いた消費者マインドの動向分析」、『日本経済2017-2018—成長力強化に向けた課題と展望—』、内閣府
- 日本経済団体連合会 [2016]「公的統計の改善に向けた提言」、日本経済団体連合会
- 柳川範之・渡辺努 [2017]「民間による経済統計の革新 —統計民営化の可能性も含めて—」、『統計』2017年1月号
- Abdulkadri, A., Evans, A. & Ash, T [2016]. “An assessment of big data for official statistics in the Caribbean,” Studies and Perspectives Series – The Caribbean No.48, ECLAC
- Bean, C [2016]. “Independent Review of UK Economic Statistics,” Cabinet Office and HM Treasury

- Bender, S [2017]. “The use of Big Data in nowcasting and forecasting of macroeconomic variables,” Deutsche Bundesbank
- Bosch, O. & Windmeijer, D [2014]. “On the Use of Internet Robots for Official Statistics,” UNECE MSIS conference, Dublin
- Cavallo, A. & Rigobon, R [2016]. “The Billion Prices Project: Using online data for measurement and research,” NBER Working Paper 22111
- Central Banking [2016]. “Big data in central banking: 2016 Survey”
- Cœuré, B [2017]. “Policy analysis with big data,” Speech at the conference on “Economic and Financial Regulation in the Era of Big Data”, organized by the Banque de France, Paris, 24 November 2017
- Daas, P. & Puts, M [2014]. “Social Media Sentiment and Consumer Confidence,” Statistical Paper Series No.5, European Central Bank
- Eurostat [2017]. “Tourism statistics: Early adopters of big data?,” Eurostat Statistical Working Papers
- Farrell, D. & Greig, F [2016]. “Paychecks, Paydays, and the Online Platform Economy,” JP Morgan Chase Institute
- Florescu, D., M. Karlberg, F. Reis, P. R. D. Castillo, M. Scaliotis, and A. Wirthmann [2014]. “Will ‘Big Data’ Transform Official Statistics?,” Eurostat
- Galbraith, J. & Tkacz, G [2013]. “Analyzing Economic Effects of September 11 and Other Extreme Events Using Debit and Payment System Data,” Canadian Public Policy, Vol. 39 (1), pp.119-134
- Glass, E [2017]. “Big data in central banks: 2017 survey,” Central Banking
- Hammer, G., Kostroch, D., Quiros, G. & STA Internal Group [2017]. “Big Data: Potential, Challenges, and Statistical Implications,” IMF Staff Discussion Note
- Henderson, V., Storeygard, A. & Weil, D [2012]. “Measuring economic growth from outer space,” American Economic Review 2012, 102(2): 994-1028
- Hoogteijling, E [2016]. “Modernisation of Price Collection at Statistics Netherlands,” ESS Modernisation Workshop, Bucharest, March 2016

- House of Commons [2016]. “The big data dilemma,” Science and Technology Committee, Fourth Report of Session 2015-16
- Hury, J. & Lamboray, C [2013]. “The use of scanner data in the Luxembourg CPI: first lessons learned,” STATEC
- IBM [2013]. “2013 IBM Annual Report”
- Irving Fisher Committee [2017]. “Big Data,” IFC Bulletin No. 44, Proceedings of the IFC Satellite Seminar on “Big Data” at the ISI Regional Statistics Conference 2017 in Bali, Indonesia, on 21 March 2017, Bank for International Settlements
- Klein, T [2017]. “Access to New Data Sources for Statistics: Business Models and Incentives for the Corporate Sector,” OECD Statistics Working Paper Series No.82
- Kroon, J [2012]. “Mobile positioning as a possible data source for international travel service statistics,” UNECE
- Landefeld, S [2014]. “Uses of Big Data for Official Statistics: Privacy, Incentives, Statistical Challenges, and other Issues,” United Nations Global Working Group on Big Data for Official Statistics
- Laney, D [2001]. “3D Data Management: Controlling Data Volume, Velocity and Variety,” META Group Research Note, 6.
- Ma, L [2014]. “A Big Data Pilot Project with Smart Meter Data (abridged version),” Proceedings of Statistics Canada Symposium 2014
- MacFeely, S [2016]. “The Continuing Evolution of Official Statistics: Some Challenges and Opportunities,” Journal of Official Statistics 32 (4)
- Naylor, J [2016]. “Big Data at ONS,” GSS News, 23 March 2016
- Naylor, J., Swier, N., Williams, S., Gask, K. & Breton, R [2015]. “ONS Big Data Project – Progress report: Qtr 1 January to March 2015,” Office for National Statistics
- ONS [2017]. “Identifying caravan homes in Zoopla data: June 2017,” ONS Methodology Working Paper Series No 11
- Sammar, M [2013]. “Past, present and future of scanner data with focus on Statistics Sweden,” Statistics Sweden

- Sammar, M., Norberg, A. & Tongur, C [2013]. “Issues on the use of scanner data in the CPI,” Paper presented to the Thirteenth Meeting of the International Working Group on Price Indices, Copenhagen, Denmark.
- Scannapieco, M., Virgillito, A., & Zardetto, D [2013]. “Placing Big Data in Official Statistics: A Big Challenge?,” NTTS Conference, 5-7 March 2013, Brussels
- Struijs, P [2016]. “BIG DATA for official statistics,” Basque Statistics Office
- Swier N [2015]. “The ONS Big Data Project,” NTTS Conference, 10-12 March 2015, Brussels
- Swier, N., Komarniczky, B. & Clapperton, B [2015]. “Using geolocated Twitter traces to infer residence and mobility,” GSS Methodology Series No.41, Office for National Statistics
- UNECE [2013]. “What does ‘Big Data’ mean for official statistics?,” Conference of European Statisticians, 10 March 2013.
- United Nations [2015]. “Deliverable 2: Revision and Future Development of the classification of Big Data,” 2015 Global Conference on Big Data for Official Statistics, Abu Dhabi, 20-22 October, 2015
- United Nations Statistical Commission [2015]. “Report of the Global Working Group on Big Data for Official Statistics,” Items for Statistical Commission Forty-sixth session 3 – 6 March 2015
- [2018]. “Report of the Global Working Group on Big Data for Official Statistics,” Items for Statistical Commission Forty-ninth session 6–9 March 2018
- Upadhyaya, S. & Kynclova, P [2017]. “Big Data – Its relevance and impact on industrial statistics,” Inclusive and Sustainable Industrial Development Working Paper Series WP 11, UNIDO
- Weber, P., Fecht, F. & Thum, S [2017]. “Capturing depositors' expectations with Google data,” IFC Bulletin No. 44, Bank for International Settlements