

ビッグデータの 統計的利活用に向けて

平成30年5月23日
統計委員会担当室



構成

1. 統計改革と基本計画
2. ビッグデータの概要
3. 今日的位置づけ
4. 国際的な取組状況
5. 利活用上の論点・課題
6. 我が国の公的統計における利活用状況

地方自治体・民間が保有するデータの利活用のための仕組み

○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等

各府省と地方自治体・民間の間における各種データの相互利活用については、現時点では一般的なルールはない。また、その推進は、個々に法令上の制約がある場合があること、偏りやノイズの程度等個々のデータの性質の違いが大きいこと、利活用のための研究主体やデータ形式の標準化・統一化の推進主体が確立していないこと等から、利活用を全般的に推進するだけでは十分ではない。

このため、ニーズが高いにもかかわらず、法制面・技術面等の課題により、利活用に至っていない各種データについて、優先度が高いものから、専門技術面も含めた関係者間の検討をオープンな形で個別的・集中的に行い、対応事例を積み重ねていくこととする。

具体的には、統計委員会が、E B P M推進委員会、各府省、地方自治体、民間等からの提案（（ア）に掲げる取組の結果、調整がつかなかったものを含む。）に基づき、利活用上の課題のある各種データの利活用について、専門技術を有する委員等及び関係者による協議会を設け、必要に応じて統計研究研修所やI C Tの専門家等の協力も得つつ、集中的に課題に対応するパイロット的な枠組を設けることとし、その具体的な内容について、年内を目途に結論を得るとともに、必要な制度・運用ルールの整備を行う。

併せて、ニーズに応じて民間データを政府統計の分類コードに基づき変換し、それを民間等に還元する仕組みも検討する。

公的統計の整備に関する基本的な計画〈平成30年3月閣議決定〉

第3 公的統計の整備に必要な事項

1 統計作成の効率化及び報告者の負担軽減

(1) 行政記録情報等及び民間企業等が保有するビッグデータ等の活用

行政記録情報等及び民間企業等が保有するビッグデータ等を統計の作成に活用することは、統計調査における報告者の負担軽減のみならず、正確で効率的な統計の作成にも寄与することから、各府省における積極的な活用が必要となっている。

一方、これらの行政記録情報等や民間企業等が保有するビッグデータ等は、①法令上の制約や電子化の状況が多様であること、②偏りやノイズなど個々のデータの性質の違いが大きいこと、③データ形式の標準化・統一化がなされていないことなどから、利用可能性の高いもの又は優先度の高いものから個別的・集中的に対応を進めていくことが重要である。

このため、総務省は、最終取りまとめにおいて、専門技術を有する委員等及び関係者による協議会を設け、集中的に課題に対応するパイロット的な枠組みを設けることとされていることも踏まえ、産官学連携による会議を開催し、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、関係者の協力を得て集中的に協議することにより利活用上の各種課題の解決や優良事例等を積み上げるとともに、ビッグデータ等の効果的な利用状況の把握に努めその情報の共有・横展開を促すことにより、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。

また、各府省は、以下の取組を通じて、行政記録情報等及び民間企業等が保有するビッグデータ等の活用の推進に取り組むとともに、それらのデータを適正に管理する。

別表：今後5年間に講ずる具体的施策

項目	具体的な措置、方策等	担当府省	実施時期
<p>1 統計作成の効率化及び報告者の負担軽減</p> <p>(1) 行政記録情報等及び民間企業等が保有するビッグデータ等の活用</p>	<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>○ 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>平成30年度(2018年度)から実施する。</p>

構成

1. 統計改革と基本計画

2. ビッグデータの概要

3. 今日的な位置づけ

4. 国際的な取組状況

5. 利活用上の論点・課題

6. 我が国の公的統計における利活用状況

2. ビッグデータの概要

- ビッグデータの「定義」
- ◆ 統一化されていない (no single definition is generally accepted)
- ◆ 文脈に応じた様々な定義づけの試みが散見される
- ◆ 時代と共に変化を続ける概念 (ever-changing concept)
- ◆ 明確な定義づけを行うことが適切ではないとの指摘も存在
- ◆ 公的統計との相違：統計作成を目的に集められたデータではない
 - “designed data”と対比するかたちで、non-traditional data, organic data, non-statistical data などの表現も多数

2. ビッグデータの概要

○ 国連・欧州経済委員会の「定義」

◆ 公的統計への活用：ビッグデータは「**データ源**」

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.”

◆ 3つのVs：「**量**」・「**(生成) 速度**」・「**多様性**」

(参考) 上記3Vsに「信頼性 (Veracity)」、「ボラティリティ (Volatility)」、「可変性 (Variability)」、「価値 (Value)」、「可視化 (Visualization)」など、様々な (Vから始まる) 特性を加えた指摘も多い。

2. ビッグデータの概要

○ 基本計画（平成30年3月閣議決定）における定義

「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」

○ 総務省「情報通信白書」（平成29年度版）による整理

1. 政府：国や地方公共団体が提供する「オープンデータ」
2. 企業：暗黙知（ノウハウ）をデジタル化・構造化したデータ
3. 企業：M2M（Machine to Machine）データ
4. 個人：個人の属性に係る「パーソナルデータ」

2. ビッグデータの概要

ソーシャル・ネットワーク情報

- SNS情報 (Facebook、Twitter等)
- ブログ、コメント情報
- 個人作成ドキュメント
- 画像データ (Instagram、Flickr等)
- 動画データ (Youtubeなど)
- インターネット検索
- 携帯電話テキストメッセージ
- ユーザー地図情報
- Eメール情報



国連・欧州経済委員会
による分類 (2013年)

IoT活用によるデータ

- ホームオートメーション
- 天候・公害センサー
- 交通情報・リアルタイムカメラ
- 科学センサー情報
- 防犯・監視カメラ映像
- 携帯位置情報
- 自動車走行データ
- 衛星画像
- コンピューターログ
- ウェブサイトログ



商業取引データ

- 診療情報
- 商取引記録
- 銀行決済・株式売買記録
- eコマース
- クレジットカード



構成

1. 統計改革と基本計画
2. ビッグデータの概要
3. 今日の位置づけ
4. 国際的な取組状況
5. 利活用上の論点・課題
6. 我が国の公的統計における利活用状況

3. 今日の位置づけ

- 統計作成を担う公的機関にとってのビッグデータの位置づけ…
- ◆ ビッグデータは、もはや “buzzword” ではない
(IMF Staff Discussion Note, 2017)
- ◆ 公的統計が射程外に置いてきた現象の把握が可能になる点で、ビッグデータは各国統計局にとって “game changer” となる可能性
(Struijs, 2016)
- ◆ ビッグデータ時代の到来により、統計局による「統計作成に係る独占的地位が終焉した」との時代認識の下、イノベーションに対応しない場合には、統計局が情報マーケットから疎外されるリスクがある
(Eurostat, 2017)

3. 今日的位置づけ

- ◆ 公的統計への活用可能性として、前掲 I M F は以下の3点を指摘
 - ・ 既存の公的統計では把握困難な現象の把握・新指標の作成
 - ・ 公的統計の公表ラグを埋めるためのタイムリーな動向把握・予測
 - ・ 新たなデータ源の発掘を通じた既存の公的統計の改善
- ◆ 他方、利活用の推進に否定的な見解も存在する点、留意の要
 - ・ 既存の統計で十分である
 - ・ ビッグデータは、活用上のリスク・不透明感が高すぎる
 - ・ 公的統計として積み上げてきた信頼や評判を失うリスク 等

構成

1. 統計改革と基本計画
2. ビッグデータの概要
3. 今日の位置づけ
- 4. 国際的な取組状況**
5. 利活用上の論点・課題
6. 我が国の公的統計における利活用状況

4. 国際的な取組状況

○ 国連・統計委員会

2014年：Global Working Group on Big Data for Official Statistics を設置

➤ 個別のテーマに基づきタスクチームを設け、具体的な検討を実施

1. Access and Partnerships
2. Big Data and the Sustainable Development Goals
3. Mobile Phone Data
4. Satellite Imagery and Geo-Spatial Data
5. Scanner Data
6. Social Media Data
7. Training, Skills and Capacity-building
8. Committee on Global Platform for Data, Services and Applications

○ 欧州統計システム（ESS）

2013年：ビッグデータと公的統計に係る覚書締結

2014年：アクションプラン+ロードマップの策定

2016年：ESS net Big Data と呼称されるデータ源に着目したワークパッケージを開始

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

4. 国際的な取組状況

○ 英国

2014年：ビッグデータ・プロジェクト を発足

➤ 4つの試行的プロジェクトを開始

- (1) ネット上の物価把握
- (2) Twitter情報
- (3) スマートメーター
- (4) 携帯電話位置情報

- ※ いずれもデータ保護に係る不安を惹起しないよう、オープンデータ・匿名化済みデータ・集計済みデータが対象
- ※ 検討の進捗は、四半期レポート (Progress report) にて公表
- ※ 従前のコンピューターでは処理できないほどの大規模なデータを取り扱うため、新たに「イノベーション・ラボ」を設立。

○ オランダ

2016年：ビッグデータ統計センターを設立

- 調査・研究のフェーズから。通常の統計作成業務への移行を企図
- ビッグデータ活用の先駆者 (global leader) となることを展望し、他国の統計局や複数の民間企業とパートナーシップを締結の上で、具体的なプロジェクトを検討・推進

(例) Google社との協働を通じたデジタルエコノミーの捕捉、Twitter情報の活用による社会センチメントの把握、携帯位置情報を活用した昼間人口の推計、等。

○ カナダ

2014年：プロジェクト第1弾としてスマートメーターの活用による電力消費量等の把握、既存統計への活用を検討 (配電事業者との協定締結、データサンプルの提供等)

構成

1. 統計改革と基本計画
2. ビッグデータの概要
3. 今日的位置づけ
4. 国際的な取組状況
- 5. 利活用上の論点・課題**
6. 我が国の公的統計における利活用状況

5. 利活用上の論点・課題

ビッグデータ活用のメリット

速報性

リアルタイムかそれに近いような速度でデータの入手が可能。

(公的統計は、調査票の配布、回収、結果の集計といったプロセスに一定の時間を要する。)

データ量の膨大さ

調査対象に限って見れば、全数もしくはそれに近い規模での把握が可能であり、詳細な内訳や属性別データの入手も可能。

低コスト

一旦データ収集の仕組みが整備されれば、データの収集に追加的なコストをほとんど要しない。

(公的統計は、データの収集のために調査票の配布、回収、結果の集計の一連のプロセスが必要で、調査のたびにコストがかかる。)

ビッグデータ活用に向けた課題

母集団代表性の担保

特定の業態など、調査対象に偏りがある場合も多く、公的統計としての母集団代表性が担保されていない。

データ形式の統一

現状では、ビッグデータの作成主体によって商品コードが異なるなど、異なるビッグデータの間でデータを集計することが難しい。

データ提供のインセンティブ付け

民間企業にとっては、営利目的で作成したビッグデータを公的統計のデータソースとして提供するインセンティブがない。

5. 利活用上の論点・課題

経済統計へのビッグデータの活用可能性・活用例

分野	活用するビッグデータ例
物 価	<ul style="list-style-type: none">• 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	<ul style="list-style-type: none">• 求人情報（リクルートなど）
消 費	<ul style="list-style-type: none">• 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	<ul style="list-style-type: none">• 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	<ul style="list-style-type: none">• 携帯電話通信データ
運 輸	<ul style="list-style-type: none">• 道路の交通量データ• 自動車の位置情報

5. 利活用上の論点・課題

- データ量
 - 精度検証のための相応のデータ量の必要
 - 時系列の指標作成・トレンド分析のための必要量の確保
- データの形式
 - 用語や様式が区々、非構造化データの取扱い等
- データ源毎の特性
 - 偏りやノイズの存在
 - 各データの発展段階の差異
- パーソナルデータにおける個人情報保護
 - 炎上リスクが顕現化した事例の存在

5. 利活用上の論点・課題

- データホルダーへのインセンティブ設計
 - 成果物の還元、統計分析のフィードバック
 - 企業のCSRと公的機関の役割
 - 対話・信頼醸成の重要性
- データ提供の継続性の確保
- プロジェクトを支える統計技術・先進的なツール
 - 公的統計と民間統計の計測ギャップの検証
 - 人工知能（AI）やウェブスクレイピング等の先進技術
- プロジェクトのコーディネーション
 - 優先順位の検討、リソース投入のオーバーラップの回避等

(参考) ヒアリングで聞かれた主な意見 (1)

※平成29年9月28日から10月13日にかけて、各種のビッグデータを取り扱う民間企業、地方公共団体にヒアリングを実施。

<データの相互利用の促進への期待>

- ◆ ビッグデータ時代にあっては、官民間問わず様々なデータを組合せることで、データの価値が一層高まる。従って、様々な主体が協力する場が必要。有識者・経験者の知見は民間ビジネスの発展にも重要であり、産官学で連携する必要がある。
- ◆ 官民間問わず、利活用のモデルケースが発掘されることで、加速度的にデータ応用が拡大する。
- ◆ 国や地方が民間データを積極的に利活用する上で、予算措置は不可欠。提供できる技術や知恵は予算規模が影響。
- ◆ 様々な原因で、国・地方が保有するデータがなかなか出てこない。「出せるものは出す」といった風土での取組を期待したい。
- ◆ 公表時期の遅れや、統計の取りまとめに当たっての算出方法・分析過程の不明瞭など、公的統計の問題点を個別に解決・改善してもらいたい。ユーザーのニーズに耳を傾けつつ、相互の利活用を図ってもらいたい。

<民間企業によるデータ提供の可能性・課題>

- ◆ 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。
- ◆ 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。
- ◆ 企業が無償でデータを提供する場合には、それが企業の宣伝となるか、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。
- ◆ 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。
- ◆ 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。

<地方公共団体の取組・要望>

- ◆ 民間のデータを活用することで地方行政の効率化を達成した事例も多い。県の政策課題の解消のために、官民で知恵を出す必要性が高まっている。特定地域における一過性のプロジェクトにとどめないためには、国の理解や支援も不可欠であろう。
- ◆ 地方等が保有する行政データに対するニーズは非常に大きい。すべてを一律にオープン化することは難しくとも、優先度の高いものからオープン化に向けて障壁を突破していくことが重要ではないか。
- ◆ 基盤データは国が整備すべきものであるが、統計の速報性も重要であるため、地方が保有する行政記録情報などを活用しつつ、公的統計の公表ラグを埋めたいとのニーズは高い。地方間で異なる行政記録情報の定義やデータの標準化などを進めるために、国の調整機能に期待。

<協議会への期待等>

- ◆ 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。
- ◆ 協議会では、民間データの提出のみをテーマとするのではなく、「双方向」でデータを提出し合う観点が大事。
- ◆ 実務に精通した少人数からなる会議体において個別に課題を解消する方が望ましい。大人数での協議会となると、調整に要するコストが大きいため迅速な問題解決には至らないと思われる。
- ◆ 国や地方が保有するデータの標準化・オープン化が進み、分析や利用の基盤となるデータの整備がなされていくことを期待。
- ◆ 技術面では民間ノウハウを活用しつつ、基盤作りやそれに対する予算を国が手当てする流れが望ましい。民間企業任せでは、独自のコンソーシアムが複数設立されるおそれ。主要なデータが分散化され、一種のデータ囲い込みが過度にビジネス化しないよう、国が間に入ってデータ整備を進めることが重要。
- ◆ 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。

構成

1. 統計改革と基本計画
2. ビッグデータの概要
3. 今日の位置づけ
4. 国際的な取組状況
5. 利活用上の論点・課題
6. 我が国の公的統計における利活用状況

6. 我が国の公的統計における利活用状況

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の概要

民間企業等が保有するビッグデータ等を活用している統計の概要			活用している民間データの概要		統計作成に活用する際の形態	活用による効果
統計名	実施機関	統計の目的	民間データの名称	民間データの収集に当たっての根拠（法令、協定内容等）		
国民経済計算（加工統計）	内閣府	国民経済計算は、2008年に国連によって勧告された国際基準（2008SNA）に基づき、一国全体のマクロの経済状況を生産、分配、支出、資本蓄積といったフロー面や資産、負債といったストック面から体系的に明らかにすることを目的とする	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	各種基礎データの一つとして活用している	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	景気動向指数は、生産、雇用など様々な経済活動での重要かつ景気に敏感に反応する指標の動きを統合することによって、景気の現状把握及び将来予測に資することを目的とする	<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系列に活用している	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	物価の変動を時系列的に測定することを目的とする	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している	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、各機種の特長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	電気工事、管工事、計装工事に関する受注高等を把握することを目的とする	<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している	調査を行わず統計の作成が可能となる