



総務省

ビッグデータ・トライアルの推進について (1)ビッグデータ・シェアリングについて

令和5年5月26日

総務省統計品質管理推進室

公的統計におけるビッグデータの活用の意義 (公的統計に関する基本的な計画 (令和5年3月閣議決定) (抄))

第3 公的統計の作成・提供・利用の基盤整備

6 国民の支持・理解が得られやすい統計作成への取組

(統計調査に関する基本的考え方)

・・・現在、統計調査の回答負担の軽減や新たな統計の作成の観点から、行政記録情報等の活用が進められる中でも、公的統計の作成の情報源は、統計調査が中心となっている。・・・

さらに、国民のプライバシー意識の高まりなどにより、統計調査の実査についても難しさが増している。オートロックマンションの増加や生活スタイルの多様化等により、調査対象者との接触が困難となっている現状に対応するためには、実査における報告者に応じた対応や配慮が従来以上に重要となっている。

統計調査を取り巻くこのような状況を踏まえると、調査実施者は、これまで以上に、統計調査の報告者の声に謙虚に耳を傾け、デジタル技術の活用などによる回答しやすい調査の実施や、行政記録情報の積極的な活用の観点から、その利用の制約への対応、ビッグデータ等の活用による調査事項の一部縮減など、報告者の負担軽減に取り組む必要がある。特にビッグデータ等の活用については、課題はあってもそれを克服する努力が民間では続けられており、行政機関においても継続的な努力が必要である。

(1) 報告者負担への配慮

(ビッグデータの活用)

統計作成において、民間企業等が保有するビッグデータを活用することにより、報告者に新たな負担を課すことなく、これまでの統計調査の一部を代替することができるようになる可能性がある。また、統計調査により作成される統計では達成し得ない早期の統計作成・公表や、これまで把握することができなかった経済社会の実態の把握ができるようになる可能性がある。

ビッグデータの活用については、データの代表性、特性の把握、継続的な入手方法の確立、体制面や技術面の整備などの課題がある。一方、データの詳細性や速報性などの面でメリットもあり、民間ではこの課題を克服する努力と試行錯誤が続けられていることから、行政においてもその活用に向けた努力が求められている。

ビッグデータの更なる活用に向けた方策

(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6月) (抄))

4. ビッグデータの活用における課題と考察

・・・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

(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、ビッグデータの特徴を把握し、安定的・継続的なデータの入手方法を確立し、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。

ビッグデータ利活用アイデアを公募するイベントの開催

- 関係団体などにおいては、ビッグデータの利活用アイデアを広く募る各種のイベントが数多く開催されているところ、公的統計への利活用に着目したイベントは見当たらない
- 公的統計へのビッグデータの利活用推進に向けて、様々なビッグデータを試験的に利用できる場を準備するため、利活用アイデアを公募するイベント（コンペ）を定期的を実施することとしてはどうか

（アイデアを募集するテーマの例）

（例1）SDG指標の開発・改良

（例2）Well-being指標の開発

（例3）KPI指標の開発・改良

（例4）その他、官民間問わず有効活用できる指標等の開発

ビッグデータ・シェアリング（試験的利用の場）

利活用アイデアの
公募イベント

データホルダー、分析者、
関係機関による共同研究

公的統計におけるビッグデータの活用

- 【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つた新たな指標の作成

イベントの成果を上げるための視点 (構成員から事前にいただいた御意見より)

スモールスタート

○初回はある程度、想定しやすいテーマとし、参加見込みの高い者（ビッグデータ・ポータル掲載企業等）も巻き込むなどにより、スキーム（成功例）を確立させることに注力することが有益

継続的な開催

○最初から良いアイデアが出てくるとは限らない。年1回でも継続的に開催し、顔が見れる場（知り合いを作れる場）を作ることが有益

コミュニティの育成

○継続的に課題解決に取り組んでいくためには、コミュニティを育てる必要がある。コンペ参加者のその後のつながりを維持してコミュニティを育成する視点が重要

利活用アイデアを募集するテーマの広さ (構成員から事前にいただいた御意見より)

テーマをある程度限定

<メリット>

- 課題を与えられることを「待っている」エンジニアも多いと考えられるため、テーマを限定することによりそうした者を動かすことが可能
- 使用できるデータや分析・研究方針がイメージしやすくなる上、審査において、提出された複数アイデアの比較検討が可能に

<デメリット>

- テーマを限定しすぎると、募集対象の縮小による優秀なアイデアの減少の可能性

テーマをある程度広めに設定

<メリット>

- 指標を開発するのであれば、例えばSDGsであれば何でもOKとするなど、テーマ範囲は広い方が多くの提案を集めることが可能（例えば、震災ビッグデータでは「震災」というテーマだけが決まっており、あとは分析者の自由な発想に任されていた）

<デメリット>

- 単にSDGsやEBPMなどでは、参加者としては何にどう手をつけて良いか分からなくなる可能性

イベントにおいて使用するデータの範囲 (構成員から事前にいただいた御意見より)

使えるデータをあらかじめ特定（参加者にビッグデータを提供）した上でハッカソンを開催

<メリット>

- データの指定により具体的なアイデア提案が期待（何のデータか未確定でアイデア募集は難しい）
- 新たなデータを利用できることをインセンティブに感じる研究者の参加が期待（データが本当に使えるかどうか定かでない中では、コンペに参加する者がいるか疑問）

<デメリット>

- 提供データの範囲が少ないと、優秀なアイデアが少なくなったり、主催者が期待するテーマ（作りたい指標）に合致しなくなったりする可能性
- データホルダーからイベントに使用するデータを得るためには、使用者の範囲を限定する必要がある可能性が高い

データを特定せず（参加者は誰でも利用可能なデータのみを使用し）アイデアソンを開催

<メリット>

- 幅広いアイデア提案が期待

<デメリット>

- 新たなデータが使えないため、具体性・新規性のあるアイデアは期待薄

⇒ まず「このようなデータがあれば、このような指標が作れる」というアイデアを募るアイデアソンを開催し、その後、必要なデータを準備し、ハッカソンを行う方法もある

その他検討を要する事項

参加者及びデータホルダー（データ提供者）の双方のインセンティブを考慮した設計

- 参加者・・・新たなデータの利活用、賞の獲得、公益への貢献など
- データホルダー・・・自社データのPR（※同業社へのデータ提供はディスインセンティブとなる可能性）

イベントのタイプ、提案成果物のタイプ

- （小）論文を提出する形式
- プレゼン形式（実開催／オンライン）

関係府省との連携、コンペ運営体制や提案成果の審査体制の構築

提案された分析アイデアの権利関係等の整理