

第19回 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10回 統計技術・データソースの多様化等検討会と同時開催)

議事概要

(開催要領)

日時：令和4年10月28日（金）16:00～18:00

場所：総務省第二庁舎6階特別会議室（Web会議併用）

(議事次第)

1 開 会

2 議 事

(1) 高松市におけるデータ利活用の取組について

(2) 消費動向に関する民間データの活用方策について

(3) 2022年度 統計関連学会連合大会報告について

3 閉 会

(配布資料)

資料1 高松市のデータ利活用型スマートシティの取組

資料2 消費動向を捉える民間データのご紹介

資料3 人流メッシュ統計データを使用した月次宿泊旅行統計調査の推計について

参考資料 第18回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9回統計技術・データソースの多様化等検討会 同時開催) 議事概要

(概要)

【(1) 高松市におけるデータ利活用の取組について】

高松市総務局デジタル推進部デジタル戦略課課長代理 細川氏から、資料1に沿って説明。主なやり取りは以下のとおり。

- こうした取組にはデータを保存するためのサーバーやそのメンテナンスも必要であり、そうしたシステムをどの程度の大きさで作るかによって深まり方が変わってくると思うが、今はどのくらいの規模感の取組をされているのか。今回お話しいただいたデータはクラウド上に置かれているのか、それとも市役所の中のサーバーに置かれているのか。また、データを取得していると過去のデータも蓄積されていくと思うが、それはどのように扱われているか。
- 高松市のデータ連携の基盤はクラウドサービスで、防災データ、レンタサイクルのGPSデータ、民間から提供されたドライブレコーダーのデータも蓄積されており、それをダッシュボードや別のアプリケーションで出力している。防災についても様々なデータの

利活用、あるいは何らかのプロセスの自動化まで出来れば良いと思う。個人的には、現在行っている推計や地図上への表示機能に対してクラウドのボリュームが大きく、フル活用できていないと感じているが、逆に使えるところがまだまだたくさんあるとも言える。また、蓄積している過去の水位、潮位等のデータについては、大学や研究所等に分析してもらえないかという考えはあるが、現状そこまではできていない。

○ 高松市で分析するのは、プラットフォームに置かれているローデータなのか、それとも高松市のダッシュボード等加工された API で出ているデータなのか。

→ 今のところ、API 出力後のものしか見ていない。

○ プラットフォーム側のデータで分析をやるのであれば、もっといろいろなことができるようになるかと思う。

○ 共通プラットフォームで様々なデータを扱えることは有意義。今後、政策の効果を検証できるようなデータが集められると更に良いと思う。政策を行う時に、何を測定すればその効果が測れるのか、成果を想定してデータを集めるとよい。また、データを近隣市町村と共同利用しているとのことで、それ自体は効率的だと思うが、防災関連データなど、全国で同じようなユースケースが見込まれるデータが全国的に統一された仕様になるなどの動きはあるのか。

→ 共通利用に関しては、高松市のルールを近隣市町村で利用しているという段階にとどまっており、それを超える共通化については、まだ意識が及んでいない。

○ 一番大変なのはルールを変えることだと思うが、実際に変えたルールがあれば教えてほしい。また、事業者も含めたエコシステムをどう回すかを考えておかないと、ただデータを取るだけで終わってしまう。継続性が大事になる取組であるので、関係者のニーズを見定めてから着手すべき。

→ どちらかという、データを活用したいという部分からスタートした部分もあるので、なかなかアウトプットにたどり着けていないところがあると思う。デジタル化も計画に乗せるところまではできたが、行政の文化として前例の壁があり、なかなか難しいというのが正直なところ。

○ 防災、福祉など、効果検証を比較的やりやすかった分野、あるいはこういうデータがあれば効果検証ができた、ということがあれば教えてほしい。

→ 防災については具体的な KPI では表しづらいが、人の非効率的な動き方についてのデータをあらかじめ入手してから動くようにすることで、現場レベルでは助かっている。福祉の分野については様々なデータが取れてはいるものの、例えばセンサー自体はデータも取れており有効であったが、そのデータをビジネス化するに当たり、データの量と質の課題もあり、なかなか進んでいない現状はある。

【(2) 消費動向に関する民間データの活用方策について】

株式会社イメージ事業開発本部先端技術部兼 株式会社イメージリサーチ事業開発部
データ利活用推進グループグループリーダー 伊藝氏から、資料2に沿って説明。主なやり取りは以下のとおり。

- 消費者データを価値あるものにしていくには、マスターの整備が重要。一方で、整備は非常に大変で、昔はどのような情報なのかを手で打ち込んでいたが、よりデータを詳細にすると数が増えていくため、大変な作業になる。技術を導入すれば楽になるのかとも思うが、マスターの整備は、昔と比べてどのように変わっているか。
- マスターの整備は非常に手のかかる作業であり、SCI、SRI+で使っている約110万アイテムの商品マスターについては、約100名の体制で整備している。商品のバーコードデータが商品マスターに一致しない時は、メーカーからいただいた情報や、POSデータ、消費者パネルデータでその商品の属性を作る必要があるため、パッケージや容量等を調べるといような地道な作業で作っている。一方、生鮮・惣菜等は販売されているチェーンによってバーコードが異なるインストアコードとなっており、商品ごとにサイズや重量が異なるなど属性の付与が難しい。最近では、生鮮・総菜の名称のテキスト情報から機械学習で分類を行う技術なども開発している。いずれにしろ、マスター整備は大変な作業であるので、可能な限り効率的に情報を集めて活用しているところ。
- 小売物価統計の家電分野の調査にはデータが利活用されているが、それ以外の分野にデータの利活用が広がっていないことは理由があるのか、それともこれから広がっていくとしているのか。また、政府側としては一企業に依存することの課題があると思うが、その辺りの継続性をどう担保するのか。
- 小売物価統計で使われている家電分野のデータは、POSデータとして網羅性が高く、悉皆調査に近いものが作られていると思う。また、家電分野は商品マスターの整備が比較的行いやすいが、食品・飲料などは全ての商品を捉えられてはならず、網羅性の部分に課題がある。しかし、食品・飲料のデータも活用可能性を検討できる余地はあると考えており、活用方法を検討しご提案していきたい。また、一企業へのデータ提供の依存は、公平性・透明性の問題、またその企業からのデータ提供が中断された時にどうするか、という問題もある。難しいところだが、多くの企業に使っていただいております。ビジネスが成立しているデータとして、社会的に利活用される、政府統計を補完できるという信頼性を獲得していきたいと考えている。
- 社会課題の状況把握において、民間のデータも併せて活用することで精度の向上、新しい切り口からの課題把握を可能にすること、課題解決を起点とした事業に結び付けて

いくというのは良いと思う。ただ、データ分析に長けている人でないと発想や切り口が見えてこない場合もあり、行政側の話を知っていると、EBPM という割にデータをうまく使えていないという事例が多い。データを使った解決に向けた発想や切り口を見つけることに長けた民間側から作って提起していくような仕組みがあると良い。また、一社のためにやっているのではないというかたちにしたければ、複数の民間企業が掛け合わせて課題を提案していけると、公平性の問題も解決するのではないか。

→ 民間側もデータを保有していても社会的な活用が難しいという難しさはあり、EBPM における計量経済学的な分析自体は当社もできないが、必要なデータとしてこのようなデータがある、このような使い方ができるというご提案はできる。各分野を得意とする関係者との連携や、データを取り扱うためのスキルやノウハウの共有も必要。民間企業や行政がもっと率直に課題をぶつけ合い、民間データも活用しつつ取り組んでいく事例を増やしていくことが必要だと考えている。

○ 仮説設定、企画部分での連携やコミュニケーションが重要。

○ 海外拠点が 11 か所あるということだが、海外のデータを組み合わせて活用すると、非常に付加価値が高いものができるのではないかという印象を持った。そのような取組を実際に行っている、もしくは今後行う予定などがあれば教えてほしい。

→ 海外拠点は、主にアジアを中心に中国、タイ、シンガポール、インドネシア、インドのほかアメリカにもある。基本的には各地に進出している日本メーカーのマーケティング支援を行っている。今回ご紹介したような網羅的なパネルデータの構築等までは進んでいないが、現地でデータを保有しているパートナー企業との連携は行っている。各地の市場動向や日本市場との違いの分析等は行っているが、日本と海外のデータとの連携はこれからの課題と認識している。

【(3) 2022年度 統計関連学会連合大会報告について】

総務省統計改革実行推進室 松井参事官補佐から、資料3に沿って説明。主なやり取りは以下のとおり。

○ 資料の「指標値」について、流動人口を全日本人人口にスケール変換したもののから国勢調査常住地人口を引いたということだが、資料の「統計値」とかなり数の大きさが違うのはなぜか。また、評価の方法で RMSPE と乖離率を出しているが、この6か月 RMSPE というのは、推定に使ったインサンプルの中の最後の6か月のデータで評価したもので、乖離率は推計に使っていないアウトオブサンプル期間の 2022 年 3 月の単月で評価されたものだと思う。一般的には、インサンプルの精度をアウトオブサンプルの精度が大幅に上回るということはあまりない。単月でなく、数ヶ月を平均して評価した方がたまたまの要素が減ると思う。

→ ご認識のとおり、人流データは国勢調査で合わせて拡大推計をしているもの。宿泊施設を抽出して差を出しているが、数字に3倍くらいの違いがあり、誤差を考えても大きいというのが率直な感想。推計結果は偶然合致することもあるので、その妥当性については継続的に評価、検証していきたい。

以 上