

第21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
議事概要

(開催要領)

日時：令和6年2月21日(水) 10:00~12:00

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) ビッグデータ・ポータル の現状と今後の取組について

(2) EBPMの推進に向けた人流データの活用について

(3) 人流データを活用した国内宿泊者数の試算について

(4) その他

3 閉 会

(配布資料)

資料1 ビッグデータ・ポータル の現状と今後の取組

資料2 EBPMの推進に向けた人流データ活用

資料3-1 人流データを活用した国内宿泊者数の試算

資料3-2 (参考) 人流データの特性について

資料4-1 「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」の開催について

資料4-2 人流データ利活用研究会の開催について(案)

参考資料 第20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11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)

【(4) その他】

議事の順番を変更し、事務局から、資料4-1及び4-2に沿って説明。資料4-2のとおり、人流データ利活用研究会の開催が承認された。

【(1) ビッグデータ・ポータル の現状と今後の取組について】

事務局から、資料1に沿って説明。主なやり取りは以下のとおり。

- 活用事例のケースが増えているのはとてもよいこと。これから事例を充実させていくに当たり、例えば、個々の企業が利益を得るためにビッグデータを使っている事例だけではなく、統計に近い使い方をしているものに絞って収集していくなどの収集方針はあ

るか。また、様々な情報ベンダーに依頼して、顧客がビッグデータを統計作成に近い使い方をした例を紹介してもらうなど、ベンダーを軸にしてやっていくのもいいのではないか。

→ 統計へのビッグデータの活用という観点は念頭に置きつつ、一般的には、活用事例が統計とどのように関連しているか分からないこともあるので、まずは世の中にどのようなビッグデータがあるかを紹介していく方針で収集していく。また、情報ベンダーにも声がけしたことがあり、ビッグデータを使った事例を多く持っている企業もある。今後連携していきたい。

○ アクセス数を見ると、正直とても少ないとは思いますが、ビッグデータを活用しようというプロ向けのサイトだということであれば、一日100人だとしても、むしろ適切な対象に情報が届いていると評価できるかもしれない。ただ、利用者がどういう人なのかはもっと分析が必要。このポータルサイトはビッグデータの利活用の促進を目指すものなので、アクセス数の増加よりも、こんなことができた、という良いエピソードを作り、このポータルから生まれたエピソードとして掲載していくことが重要。そうした取組も、進捗報告としてご紹介いただきたい。

また、法人データ、人流データへの関心が高いということであれば、そうした事例やデータを増やし、それらに関心がある人たちに向けて届けることも有効。

利活用事例ページにしろ、サンプルデータにしろ、今の掲載のやり方で本当に効果があるのかは考えた方がよい。利用者はどこに関心があるのか、インタビュー等でもう少し聞いてみてはどうか。サンプルデータは、大学のデータサイエンスの初学者に対する導入の教材として使っていきたいが、現在想定されている載せ方だとちょっと手が出にくい。例えば、統計ダッシュボードではデータのちょっとした可視化をしており、RESASやV-RESASも同様だが、そうした機能も将来的には備えても良いと思う。

資料7ページの折れ線グラフについて、ソーシャルメディアでポータルへの言及がありアクセス数が跳ね上がったということだが、ソーシャルメディアが効果的なら積極的に使うと良いと思う。総務省の公式アカウントなども見る人は見ているので、そうしたところで情報を発信すれば、より幅広い層に届くようになるのではないか。

→ SNSでの広報はまさに必要性を感じているところ。発信の仕方も含めて検討していく。

○ コストはかかるが、時事に関するデータや事例の情報を増やしていくのが最も注目を浴びやすい。特に、公的な用途に関する時事的な情報があるとよい。例えば能登半島地震の際、人が避難できていない場所、津波が来る場所などが、人流データがあればすぐに可視化して見られるようになった。こうした事例は注目度も高く、データへのアクセスや活用事例そのものも一気に増えると思う。ただ、誰が分析するのか、データベンダーが迅速にデータを出してくれるかなど分からない部分もあるので、そこはデータベン

ダーとあらかじめ相談できていると良い。

また、ビッグデータを研究で使う際は再現性が重要であり、データの時点を固定する必要があるが、現在想定されている掲載方式だと、ベンダー側がデータの中身を勝手に入れ替えてしまい、違うデータになってしまう可能性がある。世界的には DOI を付けてその時点のデータを固定するという流れであり、データの持続可能な利用という点でも、そうしたことを考慮すると良い。

→ 時事に関するコンテンツの掲載については、ポータル事業の委託業者とも協力しつつ考えたい。また、サンプルデータの提供において、データが掲載されている企業のサイトにつなぐ形だと、データの管理がどうしても企業任せになるため、再現性が課題であると認識している。データをファイル形式で提供してもらおうとしても、企業とのコミュニケーションが重要になってくるので、慎重に検討したい。

○ アクセス数やトランジションについては、民間企業の立場からすると、それがどれだけ実際のビジネスにつながったかが重要。すぐには難しいかもしれないが、ポータルに掲載することでデータが役に立った、売れたというメリットを実績として出し、他の企業からも出したい、といってもらえるという効果に結びつくまでトレースしていく必要がある。データのバラエティが充実していないと、ポータルを使う価値が下がってしまうので、データの充実という意味でも取り組むべき。

また、データの利活用によって行政コストがこれだけ下がった、こういう効果があった、という成果や効果をしっかり出していかないと、データを活用したいと思う人が広がっていかないので、1つでも2つでも成功例を作っていくことを意識した方が良い。

→ 民間企業からデータ情報や利活用事例を提供してもらおうインセンティブにもつながる話。現状はどれくらい効果があったかまではトレースできていない。ご指摘踏まえ、今後も検討したい。

○ 時事ネタを取り上げることは非常に重要。既存のデータでいいので、企業がこういうデータを持っている、ここに素材があるという情報を集めて、機動的に提供するだけでも価値がある。これは政府のデータカタログサイトにずっと要望しているのだが、なかなか実現されないで、ぜひ実行してほしい。

→ 多数のアドバイス感謝。ポータルは去年開設し、一年間試行運用を行ってとりあえずここまでこぎつけたという状態。これからは内容の充実と併せて、ポータルの利便性をどう高めるかということについても、ご指摘を踏まえつつ検討していきたい。

【(2) EBPM の推進に向けた人流データの活用について】

株式会社 unerry 鈴木取締役副社長から説明。主なやり取りは以下のとおり。

- 松島の事例で、渋滞解消の施策の効果を測ったものは、これまで他の方法で測っていた指標をデータで測ったことによりコスト削減できたということなのか、それともデータを使うことで、これまで測っていなかったものを測ったということなのか。また、データの特性として、日本の人口全体を見るのに比べ、標本に偏りはあるか。
- 宮城県としては、施策実施後の指標として渋滞解消の測定はできていなかったのも、初めてのデータ取得であった。サンプルの偏りについて、スマートフォンのアプリから情報を取っている性質上、後期高齢者にフォーカスをした分析となると難しくなる。例えば、免許返納後の移動手段評価といった相談もいただくが、流石に難しい。一方、商業エリアのまちづくりや、公共交通のリデザインなど、対象年齢が広いものについては、サンプルの偏りを気にせず取り組んでいる。

- 指標は非常に重要。データありきというよりは、社会の現状を何で測るか、というのを新しく考え出し、そのためにはこうしたデータが必要だという話ができると感じた。特に自治体などと協働する中で、これを測ろう、という指標作りの対話が重要だと思う。そこはある意味勘と経験の世界であり、競争力の源泉なのかも知れないが、立場の違う人どうしがデータの活用に向けてコミュニケーションするチャンスに見えるが、どのように取り組んでいるのか。また、unerry のデータと、携帯電話会社のデータの違いは何か。
- 指標設計については、もちろん自分たちの持っている GPS ベースのデータは何が得意かということは念頭にありつつ、自治体の方の言葉そのものを意識している。例えば、外出、賑わいといった、概念的に捉えられているキーワードはどう定義されるのか、それは指標化できるのかを一旦考えてみる、という試行回数は多いと考えている。

また、携帯電話会社によってデータのタイプは違うが、例えば基地局データは一般的にメッシュ単位等で統計加工されたデータであるため、ユーザーの細かい動きは追えないという性質をもっている。当社のデータは、活用のニーズに対して、個人特定をしないことは当然のことながら、ELSI も意識してユーザーに「気持ち悪い」という印象を与えないことも重視しつつ、どのような範囲であれば使えるかを細かく議論しながら、方法論を開発していけるのが大きな違いかと思う。

- 気持ち悪がられないようにする、というのは難しい。事業者からも、データを使える人員を限定したり、ブラックボックス型で解析したりしているという話がある。人の位置をピンポイントで追えるので、ほぼ個人を特定できてしまうが、様々な解析が可能になるという利点もある。昨今の個人情報保護の流れの中で、困っていることや苦勞などあれば教えてほしい。
- データを使ってやりたいことは非常に多くあり、こういうことに使えるのではないか、という範囲も広く、一部ではオープンデータ化が望まれていることも理解している。一

方、データは個人を特定しないという前提で取得しており、個人特定に繋がりそうな詳細な分析・可視化をしないよう社内のルールや運用で担保している。当社のデータを切り取って外に出すとその担保はできなくなるため、ローデータのオープンデータ化はためられる。このようにデータを様々な人と共有し、一緒に利活用の取組を進めるといふレバレッジを効かせづらいう状態にある。

○ 研究者もリッチなデータで解析したいが、なかなかデータを自由に触ることができず、研究がうまく進まないということもある。

○ こうしたデータは非常に有効で、ある施策を実施した場合の将来予測につながればよい。その際、属性に加え、個人の行動の意図までプロファイリングできるかというのではないかと考えている。これまで、意図をつかめた例はあるか。

→ 「意図」や「目的」を明確化するというはやれていないが、「行動傾向」をつかむことによりプロファイルをしているケースはある。例えば、ショッピングモールの中に多くのビーコンを設置し、来訪テナントや回遊性を可視化・評価するケースでは、来訪者の施設利用頻度や来訪テナントなどからプロファイリングをしている。今後、同じ回遊でも、目的意識をもっているのか、それが行動にどう表れるか、単独で行動している時とそうでないときで意志がどう変わるかということもより紐解いていきたいが、今やれることで手一杯になっているのが現状。

○ 指標のモニタリングはぜひ進めていくべきだが、業界団体の中で、データホルダーごとの弱いところを補完しようとする、そのデータをどう扱えば良いか分からなかったり、コストがかかったり、団体の運営が難しかったりすることがあるかと思う。この点に関して、行政への要望はあるか。

→ 業界団体の設立直後から、位置情報活用ビジネスの健全な発展のため、データプライバシーに関するガイドライン作りをやってきた。法律で規制されているものに対して、どこまでプラスαでやっていくか、企業体力の違いなど様々な観点を考慮しながら検討しており、また、個人情報保護委員会含め、複数の関係機関からアドバイスをいただいたと聞いている。

業界団体として、データの利活用範囲を拡げ、様々な社会の課題解決に寄与していきたい、というのが共通の思いであり、そこに向け、私個人として現時点で行政への具体的な要望があるわけではないが、今後も、行政と連携して、健全性データ利活用を進めていきたい。

【(3) 人流データを活用した国内宿泊者数の試算について】

事務局から、資料3-1及び3-2に沿って説明。主なやり取りは以下のとおり。

- データの比較は非常に重要である。会社によって値が大きく異なっており、公的統計調査の結果と比べると、地域特性による傾向の違いがあるが、それがなかなか一般には共有されていない。公的統計を用いてデータを検証し、その比較を強調したものを出した方がいいのでは、と強く思った。ベンダーは、データは自社で調整しているので問題ないと言っているが、実際に解析してみるとやはり違いがみられる。ビッグデータ・ポータル上でもよいが、各府省庁がデータの特徴を参照できるよう、知見を広めるべきである。
- 各社のビッグデータの違いはデータの取得法の違い、推計方式の違いに由来すると考えられるが、業界のバランスを考慮すると、公的に如実に言い過ぎないように配慮することが必要。行政として、オフィシャルなガイドラインを示していくことを検討していかないと、ユーザーのデータ利活用の利便性を高めるつもりが逆効果となり、ユーザーのデータに対する不信感を助長したり、メッシュ統計データ販売業者の事業を阻害したりするだけの結果になってしまうと思う。
- 資料3-2について、常住人口に合わせるために観測した人数を膨らませているため精度が低くなるということで、各社のデータを合わせる工夫が必要だと感じた。また、資料3-1について、宿泊施設の位置するメッシュはどのようにピックアップしたのか。宿泊者数のリストが別にあって、その宿泊施設があるメッシュを取ってきたとすると、宿泊施設の規模感を踏まえて重み付けをすれば精度が上がるのではないか。また、モデル選択の部分と評価する部分で、評価する指標をあえて変えている理由は何か。
 試算はなかなかうまく当たらないという話だが、利用者としては、方向性だけでも当たると嬉しいケースもあるので、方向性が当たっているかを評価するということもありえる。また、今回は目的が違うかも知れないが、メッシュデータでもオリジンとデステイネーションがあるメッシュデータが使えると、もっと正確に予測できるのではないか。
- 宿泊者数リストは、OpenStreetMapから宿泊施設の位置情報を把握し、それをメッシュに変換している。ただ、部屋数まで分かる網羅的なリストではないので、重み付けまではやっておらず、今後の課題として検討していく。なお、モデル選択は、本来的には検証データを用意して総合的にやるべきものだが、期間が短いため、観測データのみで当てはまりがいいものを使っている。
- データを比較することは、ビッグデータ・ポータルに様々な活用方法が集まることにもつながり、有意義である。例えば医学分野だと、薬と症状と処方との組み合わせを集めることで、どういう薬ではどういう症状に対してどういう使い方をすればいいのかを統合的に調べるといったメタ分析があるが、それに近いといえる。データ×目的×分析手法のマトリックスが多く集まることで、ビッグデータの活用につながる。公開の仕方は考える必要があるが、そうした分析につながると素晴らしいと思う。

- 正確な国内宿泊者数を割り出すことは容易ではないが、各社のデータの違いを踏まえ、て宿泊者数の推計を行っていくことで、実際は働いている人、通過しただけの人などの情報を合わせていくと、本当の夜間人口を割り出すことはできるのではないだろうか。そのような結果は防災などの分野にも活用できると思う。

- データソース間のばらつきを示したことは非常に有意義である。データの用途は様々だが、例えばまちづくりに活用する際、データのバイアスが意思決定に副作用をもたらす可能性もあり、そうした怖さも含めて具体例を示す点で有意義な分析だと思う。

以 上