

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と同時開催)

議事概要

(開催要領)

日時：令和5年5月26日（金）16:00～18:15

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・シェアリングについて
- (2) 栃木県 産業データの共有と利活用について
- (3) ビッグデータ×情報デザインによる社会課題解決
- (4) ビッグデータ・ポータル開設について

3 閉 会

(配布資料)

資料1-1 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

資料1-2 コンペ開催形式案の一覧

資料2 産業データ共有・利活用プラットフォーム事業 令和4年度実証実験について

資料3 ビッグデータ・トライアルの推進について

- (4) ビッグデータ・ポータル開設について

参考資料 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 同時開催) 議事概要 (概要)

【(1) ビッグデータ・シェアリングについて】

事務局から、資料1-1及び1-2に沿って説明。主なやり取りは以下のとおり。

→ データソンやハッカソンの開催は非常にいい取組。これらにはいくつかのタイプがあるが、ポイントとなるのは開催期間。自由度を高めにするデータ分析したり、アイデアから考えたりしなければならないので、開催期間を長めに取る必要がある。国際会議でもデータソンやアイデアソンがあるが、大体丸3日は使い、データに関する議論を行った後にチームを組み、学会最終日の午前中で発表し、評価、表彰という流れ。これくらいの短い期間であれば、データやテーマをしっかりと決めて出すという形になる。

- どのような形式でもよいが、想定している期間によってデータソンかハッカソンか、データやテーマを広く設定するのか限定するのか、変わってくると思う。
- 初めての取組であるので、リソースも考慮しながら、出来ることから小規模でやるというイメージでいる。また、コンペの形式をどうするのかは、どの程度の完成度のものを目指すのかにもよると思うが、コンペで出てきたものは共同研究を通して公的統計に使えるものにしていくイメージであり、まずはその前段階のアイデアを募るところにフォーカスしている。
- 国際会議でのイベントでは、若手の研究者や学生がメインで作成し、表彰された良い作品について共同研究が始まることもある。それくらいの規模感で進めるのが良いと思う。
 - 参加者のインセンティブとしては公益への貢献になると思うが、イベントを開催する際には、その統計指標を使いそうな関係者に審査員に入ってもらおうと良い。デジタル田園都市国家構想に取り組んでいる自治体の関係者などに、自分が使うとすればこういう指標を使いたい、というように評価してもらおうと、自分の作成した指標は公益のために使われるんだ、というインセンティブになる。指標を使う側も一緒になって良いものを見つけていくことが大事。
 - 参加者のターゲットによってアプローチの仕方が変わってくる。データホルダーや、企業内のデータの活用をしていきたいというステークホルダーを呼ぶのであれば、ある程度テーマを絞る必要がある。誰にどういう目的で参加してもらおうかによって、テーマ設定、呼びかけ方、広報の仕方も変わるので、ターゲティングをすると良い。
 - 東日本大震災後のビッグデータワークショップでは、2012 年で何が起こったのか明らかにすることを最大のモチベーションに、様々なデータホルダーがいろいろなデータを出し、研究者も 500 人程集まって、オールジャパンで取り組んだ。あれから 10 年経ったが、コロナが一体どのような爪痕を残したのかはまだ誰も検証していないので、こうした場をうまく使って検証するべきである。SDGs 関連に取り組むのも良いが、コロナ対策には巨額の費用がかかっており、震災の復興費用より規模が大きい。いろいろなビッグデータをテーマに、たくさんの人に論文を書いてもらったら良いのではないか。また、ビッグデータを使う際は、AI をどう使うのかは必ず議論に入れなければならないと思う。
 - 公的統計への活用という前提がある以上、参加者への動機付けは工夫する必要がある。知りたいテーマがたくさんあるのに対して、使えるデータが少ない現状を踏まえると、データホルダーのモチベーションが前提としてまず重要となる。SDG や Well-being の

指標開発をしたいというくらいの広めの設定にして、賛同したデータホルダーが出してくれたデータの中でテーマを広めに設定するのが良い。参加者として学生、研究者、企業などどこを想定するかによって、やり方は変わってくる。

- 一般論として、ナウキャストという明確な目標であればデータホルダーにとってもよく分かる一方、新指標の作成となると別次元の話になり、政策においてどういう形で使われるのかが、外部の人には分かりづらい。新指標の開発をテーマにするのであれば、政府の側で、こういう政策を推進したいが、こういうデータが足りないなどといった詳細なディレクションを示す必要がある。
- 「既存の公的統計では捉えることのできなかった新たな指標の作成」の中に、ナウキャストの指標も含んではいる。

【(2) 栃木県 産業データの共有と利活用について】

栃木県産業労働観光部産業政策課次世代産業創造室主査 狩野氏から、資料2に沿って説明。主なやり取りは以下のとおり。

- 企業の背中を押すような、素晴らしい取組。一方、こうした取組は本来各企業の企業努力でやっていくべきであるところ、競合他社の分析もやるとなると、その地域の企業からシェアをどう奪うかという話にもつながる可能性がある。この点について、企業から批判が出ずにうまく回す方法があれば他の地域でもやっていくべきだと思うが、成功要因を教えていただきたい。
また、伴走する際にはセグメンテーションを分けることで、行動パターンやライフスタイルなど多様な分析が出来るので、いろいろなアクションが可能になる。そうした部分で初めの一歩が踏み出せない企業の支援は、事業会社とうまくタッグを組んでやっていけるとよい。
- 他の商業施設の比較については、そうした分析をして大丈夫なのか、という話は企業からもあった。今回使った携帯キャリア会社の人流データは、誰でも使える形で販売されており、分析側で恣意的にデータを抽出したものでないという整理のもと、誰でも見られる結果として掲載したことが成功要因の一つだと思う。また、セグメンテーション分けについてはおっしゃるとおり、アンケート調査を行ったところ、ライフスタイルを初めとする約20項目のアンケート回答データが取得できた。そのデータを分析していく中で、商品ラインナップや売り場について、こうした方がより良いのではないかという議論はいろいろあったが、東武宇都宮百貨店の企業ブランドや方針を踏まえつつ、どこまで提言するかは検討していたところ。
- 次のステップとして、宇都宮駅前の経済圏全体で協調し、イベントをやっていくなど

の動きにつながればよい。また、人流データは初期導入のコストが問題になっていることだが、九州の自治体の例で、人流データを販売している会社と交渉して購入したデータを、民間企業に必要な分だけ安く提供しているという話もある。プラットフォームについても検討されているとのことだが、場合によってはデータをまとめて購入して、必要な分を効率的に提供するような機能もあった方が良くもしい。

→ 店舗のみならず、いずれはエリア単位や市単位等、大きなエリアでも分析していきたい。また、二次提供の件については一度交渉したものの、プラスのオプション費用が必要という話になり、悩ましいところであった。成功している自治体にヒアリングもしつつ、取組を検討したい。

○ 人流データ、POS データは比較的広く使われているので、販売者側も知見がある。ただ、一般的にデータというのは使用側がいろいろな使い方をするので、データホルダー側も価格を付けられないうえ、契約書をどう作ればよいか、他社がどう契約を結んでいるのかについての知見がないため、大変な作業になる。例えば、データの利用目的等を定めることでリミテーションがかかり、使用側のコストもかなり低くなり、ベンダー側の費用もかからないというように、どういう契約が可能なのかという点において、地域のデータを使用したい企業とベンダーの契約をつなぐサービスがあれば良いと思う。

→ 契約部分はまだ意識が薄かったところ。ベンダー側と使用側両方の立場から、どういった契約を結ぶのが良いか、知恵を絞っていきたい。

○ この手の分析は、アプリのデータと連結して ID-POS を構築するのが王道であるので、ぜひ引き続き取り組まれると良いと思う。また、次のステップとして、施策の効果検証の分析等をするとう分かりやすい。

データ提供のためのプラットフォームは時期尚早という話もあったが、例えば ID-POS などは企業からすると機密情報のようなもので、提供は難しいと思う。人流データ以外で、どういったデータがプラットフォームで提供可能だろうという話になったのか。

→ 効果検証については、今後複合的な視点を持って分析の取組を行っていきたい。また、プラットフォーム構築に当たっては、行政のデータをどのように活用するかについて、栃木県の統計課、行政改革 ICT 推進課等も含め相談していた背景がある。まず、統計課等を出しているオープンデータについては、プラットフォームに無償で掲載してどんどん使ってもらうことが望ましく、そうしたところは充実させたいと考えている。一方、現在出している統計データを民間企業が使う場合、どのような種類のデータが必要で、どのような形式が使いやすいのかについては、いろいろと悩みや課題もあり、議論を重ねているところ。スモールスタートの段階ではあるが、今後ホームページ等で情報を展開していく中で、オープンデータについては、こうした使い方をすることでこういうことが出来た、という事例を横展開していきたい。

- まずはデータを集めて、傾向を見るという大きな一歩を踏み出すサポートをしたという理解だが、次のステップとして仮説を立てて検証し、施策を改善するというサイクルを回すことが、企業の取組を促すと思う。また、目的によってデータの取り方、使うデータが変わるので、そのあたりのメニューが増えていくとユーザーにとっても良い。
- 横展開に当たっては、単に事例だけではなく、こういう使い方ができるという情報も併せて提示していきたい。

- 誰がどこから来てどういう風にお金を使うのか、という情報が全てつながっていることが一番重要。そういう意味では ID-POS が一番強く、それがアプリと結びついていれば、アンケート等で、事前にどういう情報を受け取ったのかも分かる。既に議論があったと思うが、ID-POS 系もいろいろな形のデータがあり、国としてそうしたデータをどれくらい扱うか、データホルダーが傷つかない形でどの程度公共目的に使えるのか、という部分についても議論が進むと良い。
- データを自分たちで収集するのか購入するのか、その際のコスト等も含めて選択肢をたくさん持ちながら、企業に提示していきたい。

【(3) ビッグデータ×情報デザインによる社会課題解決】

ネブラ株式会社 CEO 阿部氏から説明。主なやり取りは以下のとおり。

- 世界で見るとビジュアル化に関する研究は多いが、日本はあまりそうした研究をしない。ビッグデータのビジュアル化は、スモールデータのビジュアル化と比べ、そもそも背後のシステムが違うのに加え、大きなデータを高速に動かす仕組みも必要。どういったところが大変だったか。
- 2013 年の段階で、膨大なデータを可視化するためのプロジェクトを立ち上げた経験がある。当時は、ハリウッ드의 CG 映像の作成と同じ要領で、パソコンを 200 台程繋げて、人が動く様子を再現したほか、データ分析も卓上スパコンを置いてやっていた。それが最初の 1、2 年だが、現在はクラウドのデータベースを利用しており、小さなパソコンでも動かすことができる。当時は、ディレクター・記者・プログラマー・デザイナーの混成チームで取り組んだ。ハッカソン等にも言えることだが、アイデアを出す人と作る人が一緒にいることが必要。なお、記者はプログラミング、プログラマーは取材をするように役割をスイッチさせながら進めると、双方のスキルを使いこなせるハイブリッドな人材育成につながる。今回の国の取組でも、学生や若手研究者を対象に、そうした取組をやっても面白いのではないか。

- 因果推論、予測モデルなどの前にデータを可視化することは重要。90 万種類のデータをリアルタイムで収集しているとのことで、API が用意されていれば簡単にできるとは

思うが、リアルタイムで取れるようになっていないデータがほとんどだと思う。どうやって実現しているのか。

→ API で公開されている国の統計データは即取得するようにしており、API がなかった当時はスクレイピングで取得していた。Excel 形式のデータもタブごとに、表形式のデータも含めて全てばらしてデータ化しており、PDF も同様に取得していた。例えばトラックのデータなどは、地震が発生した日のうちにデータを保有している会社と交渉して契約し、データの仕様をもらって取得、可視化、人や車の動きを可視化するところまでを6時間くらいでやる。そこからは継続してデータを取得するので、3、4日同じデータを使いながら、物流や道路の復旧状況の分析をするくらいの速度感でやっている。ただ、やっているメンバーは2、3人。研究も企業もそうだが、突き抜けた人をうまく育てるとよい。コロナが発生してから、コロナに関するあらゆる論文を自動的に収集する仕組みを作ったが、これもそうしたメンバーで行った。

○ これまでの報告でもデータが高価であるという話があったが、データジャーナリズムとしては、基本的にはデータを購入し続けていくのか。

→ データを持っている会社であっても、データの価値の開拓が出来ていないケースがある。データを出していない企業でも、そのデータを可視化するとういうことが分かるということを取引の最初に示し、最初の事例になれば、全面的に協力が得られることがある。お互い Win-win になれる条件があれば、まずデータをもらえることもあり、データの価値が大きくなれば、そのデータをいろいろな人に売ることによってビジネスにもなる。データは10人にしか売れないとなれば高いままだが、1万社に売れるとなれば安くなり、流通していく。

人流データは、2013年にテストとして使い始め、それからは定常的に購入している。人流データによって、津波発生時ほどの浸水域に何人いるのかが分かり、ヘリコプターを飛ばすことが出来たり、原発に障害が発生しても、30キロ以内に何人いるのかすぐ分かるといいう仕組みを作ることができる。同様の仕組みによって、コロナの時には、渋谷の繁華街に何人いるのかということも、いつでも分かる状態だった。

○ 関係する研究分野としてはインフォグラフィックスがあると思うが、残念ながら日本はこの部分に弱いところ、現場でこれだけ取組が進んでいるのは非常に心強い。今後、90万種類のデータに対してある観点を与えることによって、自動的にニュース原稿を作成したり、それをAIキャスターが読み上げたり、自動的に再現ビデオをCGで作ったりなどということは考えているのか。

→ 90万種類のビッグデータを把握した上で政策立案が出来る存在として、政策立案AIなどがあるが、まさにそれに近いものを出すべくやっているところ。ビッグデータ時代は終わったという有識者も出てきているが、ビッグデータを使ってどのように情報を再

構成し、要約して、尖ったところを言語化するのが重要。見え方のマッピングだけではなく、何が情報として重要なのかをデザインしていく部分については、ジェネレーティブAIはかなり使えるのではないかと思う。

【(4) ビッグデータ・ポータル開設について】

事務局から、資料3に沿って説明。主なやり取りは以下のとおり。

○ サンプルデータの提供に当たっては、様々なビッグデータのメタデータをたくさん集めていただくと良い。その際、検索の仕方も工夫して、自然言語処理で連想検索のようなこともできると、目的に応じたビッグデータを探しやすい。

○ 検索性を高めるためにも、これからデータの掲載数を増やして価値を高めていかなければならない。いつまでにこれくらい、という小目標を積み上げていく必要があると思うが、そのあたりの想定はあるか。

→ 具体的な数値目標はまだ立っていないところ、SNSを活用した広報活動を行い、掲載希望を募ったり、連携会議に登壇いただいた関係者から御紹介いただいたりなど、草の根的に増やしていくことを考えている。

また、現在は試行的に運用しているが、具体的にどのような状態になれば本格運用に移行できるのか、またコンテンツの件数をどれくらいまで増やすのか、また、それが今の運営体制でどの程度実現可能なのかということについて、並行して検討していかなければ、現実的でないと考えている。折を見て御報告したいので、アイデアや御意見をいただければと考えている。

→ ビッグデータ・ポータルはまだまだ試行運用中のところ、地方公共団体も似たような状況があると思うが、各省の政策当局の幹部が文系出身で、なかなかデータ利活用に関心を持ってもらえないという状況もある。そうしたところにビッグデータ・ポータルを見てもらい、自分たちもこういうことが出来るんじゃないか、と思ってもらうことが重要なので、出来るだけ参考になる事例を揃え、活用してもらいたいと考えている。こうした取組の推進においては、もちろん最先端の利活用事例も必要だが、一方でEBPMの取組を各政策現場でそれなりに出来るようにするという底上げの取組も必要。その両面、特に底上げの方を担っていきたいというのが政策当局の思いとしてある。

○ こうしたコンテンツは、足で稼いでいく必要がある。勘と経験ではなくて、データドリブンに、という価値観は重要。

○ この手のデータの活用が進むかどうかは、突き抜けた事例が出るかどうかには尽きる。情報を探すのもよいが、こどもビッグデータ、介護高齢者ビッグデータ、観光ビッグデータなどについて、閲覧した人がそのデータセットをそのまま使おう、となるような突

き抜けた活用方法を見せてほしい。連携会議の関係者がグループになり、研究者や学生を使って、様々なビッグデータを使った分析の成果を示すようなイベントを開催すると、強い推進力になる。特にこども関係のビッグデータについては、ニュースで取り上げやすいものでもあり、取組を検討すべき。一見取れないようなデータでも、他のデータで代替できるのがビッグデータの面白いところであり、そうした置き換えの視点を持つと、こども関係のビッグデータはかなりあるのではないか。こども家庭庁とも連携し、何らかの取組の成果を2、3ヶ月で出したり出来ると良いと思った。

以 上