

第9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

1 日 時 令和2年2月26(水) 15:00～17:00

2 場 所 総務省第二庁舎 6階特別会議室

3 出席者

- ・ 構成員 高橋座長、庄司構成員、田原構成員、水野構成員
佐藤構成員（オブザーバー）
- ・ 審議協力者 神戸市都市局・企画調整局
ヤフー株式会社
- ・ 事務局 総務省政策統括官（統計基準担当）付 統計委員会担当室

4 議 題

- (1) ヤフーが提供するデータソリューションサービス
- (2) 神戸市におけるビッグデータ活用の取組

5 配付資料

資料1 データソリューションサービスのご紹介

資料2 都心三宮の再整備について

資料3 都心三宮再整備事業一覧

資料4 産学連携したデータ活用の取り組み

参考1 第8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

6 議事概要

(1) ヤフーが提供するデータソリューションサービス

- 議題1に関して、ヤフー株式会社の阪上様より、資料1に即して、ヤフーにおけるデータソリューションサービスに関する説明及び同社が提供しているデスクリサーチツール（DS. INSIGHT）のデモ紹介が行われた。

概要は以下の通り。

- ヤフーは多くのビッグデータを保有しており、民間企業のマーケティングや公的機関の施策とコンサルティングを含め、様々なところで活用されている。
- 自治体については、ヤフーが持っているデータと自治体を持っているデータを掛け合わせて地域支援を行っている。

主な質問・意見は次のとおり。

- ヤフーにおいてビッグデータ人材の充足状況はどうか。不足しているならば、特に足りないと思われる分野はどこか。また、内部のみならず外部の人材を活用することを考える必要があるが、外部の人がデータを扱える環境にあるのか。また、どのようなバックグラウンドの人が必要か。社会に広く伝えるためにはデータ分析ができるだけでなく、データを理解した上でビジュアル化のできるデザイナーのスキルも有する人材が必要と思うがどうか。(構成員)
 - ▶ ビッグデータ人材に関しては、ヤフーは多くの人材を擁するが、昨年秋に本格化したこのビジネスが今後さらに拡大すると厳しくなる可能性はある。ただ全員がすべての分野においてのトップ人材である必要は必ずしもなく、社内的には人材を階層化して、①トップコンピューターサイエンティスト(ヤフー研究所の研究者など)がデータ活用のフレームワークを構築、②アプライドエンジニアがそれに沿ったツールやアプリケーションを開発、③顧客担当アナリストがそれらのツールを使って現場でコンサルティングと、ある程度役割分担して進めている。社外でのデータ活用としては、大学の研究室などと契約を結んで、共同研究をクローズドな形で実施することはあるが、現状ではデータを外部で利用していくベースは基本的にはない。分析、集計、ビジュアル化も含めて全て一人で実施する必要は必ずしもなく、協業体制が重要。各特定の分野からというよりは自分の得意分野を軸足に、コミュニケーションを取っていくことが有効であり、そこをつなぐことができる人材が今後より必要になると考える。(ヤフー)
- ビッグデータとして、どのようなものを保有しているか。また分析にあたってはデータを細かくしすぎないでマクロに捉えた方が有効との説明があったが、具体的にはどういうケースか。(構成員)
 - ▶ ビッグデータの種類としては位置や検索の情報だけでなく、デモグラフィック情報やカーナビ、乗り換え検索などがあり、これを個人が特定されないよう配慮しながらヤフーID でつなげて分析に使用している。位置データのメッシュは細かすぎると意味を読み取ることが難しくなるので、分析のテクニックとして、集計エリアを適切な大きさにして、居住市町村とか、回遊パターンとか別の切り口で定量的に分析することで、付加価値の高い情報を提供している。(ヤフー)

- データアカデミーでは、シナリオの設計とか、課題設定が大事とのことであるが、どのように教えているか。(構成員)
 - ▶ 最初にヤフー側で用意したデータと課題を使用して、ワークショップ形式で実際に分析やプレゼンを実施してもらい、次にそれぞれが持っている実課題で、どんなデータが必要かということも含めて分析してもらおうというのが基本的なやり方である。(ヤフー)

- 位置情報に関して、携帯キャリア等と比べてヤフーのデータを選んでもらえる理由は何だと考えているか。ヤフーの ID を持つユーザーが、検索履歴など自己のデータを見ることはできるのか。(構成員)
 - ▶ データそのものに加えてデータをどう活用するかを実課題に即して一緒に検討してきたプロセスを評価いただけていると考えている。またお客様からの自己データに関する開示請求フローを用意しており、使用して欲しくないデータがある場合は当該データを除くなどの非開示要求に対応することが可能である。(ヤフー)

- 公的データとヤフーデータを組み合わせた事例やそれぞれのどのような点に利点を感じているかなどについて教えていただきたい。(構成員)
 - ▶ ヤフーのデータは全量のデータではないが、公的データは全量なので高品質であり、質の良い宝の山があるという印象。一方で、提供間隔が長くなるという特性があるが、ヤフーのデータは即時性がある。これらの情報を組み合わせて、情報をリッチ化して補完的に使用することで有効活用ができるものと考えている。(ヤフー)

- 公的統計の実施にあたっては、調査対象となる国民や企業に調査協力をお願いすることになるが、回答率を高めるための事前アナウンスの時期や方法に関して、ヤフーのデータを利用して効率的な方法を探ることはできるか。(構成員)
 - ▶ イベント毎に異なると思うが、仮説を立てた上でキーワード検索をかけて件数の推移を見て適切なアナウンス時期を推定することは可能だと考える。別の視点では、観光分野でも活用している手法だが、PR するときに著名人を登用するなど、効果が高い方法を検討することも可能だと考える。(ヤフー)

- 検索サービスは幾つかの会社が提供していると思うが、性別・年齢などデータの偏りはどうなっているか。またデータの組み合わせとして欲しいデータはあるか。(構成員)
 - ▶ ヤフーのアプリに関しては、スマホを日常的に使う日本人の会社員等の現役世代が中心である。分析においては、一定の年齢以下の年少者のデータは使っていない。全体としては若者～年配者までなだらかに分布しており、大きな偏りはないと考えている。欲しいデータとしては、地域のニーズとしても最終的

に経済振興が重要になることもあり、商業に関するデータが考えられる。(ヤフー)

- 公的統計を考えていく上では、データの安定性が重要であるが、ヤフーのデータのボリューム感と安定度はどうなっているか。また、DS. INSIGHT のリアルタイム性はどうか。(構成員、総務省)

- ▶ 月とか週とか日で大きくぶれることはあまりないが、法律改正やGDPR (一般データ保護規則) など政策に対して大きな影響を受けやすく、データ構造が変わる可能性はある。DS. INSIGHT のタイムラグとしては、検索データは3日前、人流データは前日までのデータが反映されている。(ヤフー)

- 年間の検索キーワード数が約92億ワードとあるが、日本全体の中での割合と、ヤフーIDでログインしたユーザーで検索している人の割合はどれくらいか。また位置情報については、どのアプリから何人くらいのデータを基に作っているか。(傍聴者)

- ▶ 92億ワードの中身はロングテールの様相を呈しており、テールには多くの意味をなさない文字列が含まれている、ツールではこれらのテールワードを足切りした上で多くの方々に役に立つワードを使用しているが、その部分は日本全体のニーズをあらわしていると考えている。検索クエリのシェアやヤフーIDでの検索の割合は非公開であるが、アマゾンや楽天のサイト内の検索を除き、グーグルとヤフーで市場のほとんどを占めている。また年間デジタルリーチ(人口のうちの推定到達率)ではグーグルで56%、ヤフーで54%ぐらいあり、十分なボリュームがあると考えている。位置情報については提供を許可したユーザーのデータから作成しており、アプリの人数は非公開であるが、ヤフー全体としては月間のアクティブなIDが約5,000万である。(ヤフー)

- DS. INSIGHTによりオレオレ詐欺で騙されやすい人が多くいるエリアを検索する等、悪用されるという可能性に関して、社内で議論しているならば、教えて欲しい。(構成員)

- ▶ 悪用の可能性があることを想定した対応として、ツールの申し込み者に対して審査を行い一定の信用のある組織体のみを提供するというを基本としている。ツールとしてはデータの大量一括ダウンロードを不可とし、また、利用したログを取得・保管しておき、何か起きた時に事後に追跡できるようにしている。(ヤフー)

(2) 神戸市におけるビッグデータ活用の取組

- 議題2に関して、神戸市都市局の清水様より資料2・3に関して、神戸市企画調整局の松崎様より資料4に関して説明が行われた。

概要は以下の通り。

- 神戸市では三宮再整備において、ビッグデータを活用して検討を進めている。プロジェクトにおいて、①現状分析、課題抽出を行って戦略を練る段階、②設計・デザインを行う段階、③その過程で実施する社会実験、④整備後の効果検証など、いろいろな場面で活用が期待できる。
- また、三宮再整備のみならずスポーツや消防など幅広い分野においてビッグデータが活用されており、職員の人材育成、市民や地域の事業者への啓蒙なども含め、積極的に取組が進められている。

主な質問・意見は次のとおり。

- 神戸市の取組は先進的なものだと思うが、なぜそれが可能になったのか、他の自治体と何が違うのか。(構成員)
 - ▶ データ活用を先進的に進めていたサンフランシスコの事例を参考に、成功事例を積み上げることにより住民や職員の意識を啓蒙し、データ整備及び事業の成果を出すことのサイクルを回していくことで、そのサイクルに民間パートナーも参加してきたという形である。一般的な自治体では単発の実証実験としてデータ活用を行っているケースが多いが、神戸市ではそれぞれの実証実験が全体政策の中でどう位置付けられているかを明確にしながら進めており、施策的にもパートナーやデータの種類に関しても、幅広いトータルな取組として行っているものである。(神戸市)
- これまでの産官学連携というと、学は大学というイメージが強かったが、今回の事例で、高校においてデータ活用の事例を示したことは、他地域への展開を図るという視点で大変価値があると感じているが、このような方法を採用したきっかけを教えて欲しい。(構成員)
 - ▶ 産官学と言ったときに、該当する学校現場と企業が近い立ち位置にいたからであり、高校が市立か県立かなど特別な理由があったわけではないと考えている。これまで企業においても、プロや大学相手にしていたことはあっても、高校を相手にしたことは無かったが、実際にやってみると、該当の高校でデータ分析が実施できることが実証され、生徒や先生の意識啓蒙やスキル向上がはかれ、その結果として、該当の高校は科学技術高校と名前を変え、該当企業への就職や、筑波大学への推薦入学が実現するという効果もあった。(神戸市)
- 「三宮クロススクエア」の社会実験において、車線が減ることが、事前に

SNS で発信されたり、カーナビのルート検索に反映されたりという形でサイバー空間で共有され、それがフィジカル空間での行動に影響を与えるといった相互作用に関しては、検討したのか。(構成員)

▶ 今回の社会実験では、「迂回してください」等の事前アナウンスは、あえて行っておらず、1 カ月間の長い実験期間をもうけることにより、ドライバー自身が学習して自然に迂回することに任せていた。本番は第一段階が 2025 年頃になるが、カーナビからのルート案内なども含めて、どうマネジメントするかを考えていきたい。(神戸市)

○ 神戸市の取組は、流行のデータサイエンスを駆使するというよりは、「課題設定→データ分析→グラフ・マップで可視化」という基本的な方法論を使用して、一定レベルの能力を多くの人に身に付けてもらうことを主目的にしていると理解した。逆にデータサイエンティストでなければできないような取組も行っているか。(構成員)

▶ EBPM 推進体制の構築の一貫としてデータサイエンティストを条件付きの職員として採用し、兵庫県との合同庁舎建設後の人流分析やタワーマンションによる地域の変動予測などを実施したことがあるが、分析を進めると例えば税データは税部局の人しか見られないといった個人データの共有という壁にあたるのが常であり、データ所管組織に兼務して分析するとか、限定した形でしか対応できず、思ったような取り組みができなかった。今後、個人情報保護法改正のいきさつなども勉強しながら、データ管理のルールを作成する等の課題の検討が必要。(神戸市)

○ ビッグデータを活用するにあたり、データ購入やセンサー取り付け費用の確保や、職員の時間も必要になるが、長期的に見たときに、費用対効果はどのようになるのか。(傍聴者)

▶ 取り組み自体が業務効率の向上につながっており、特にGISでは紙で行っていた業務が現場でもタブレットが使用されるようになり、肌感覚では確実に効率化がはかれていると感じており、今後これを可視化していく必要があると考えている。(神戸市)

○ 人流センサーや「AI-Smart 空調」のデータは防災などに有効だと思うが、その情報は設置した企業以外に開示されるのか。また、今後センサーを増やすとした時に、どのような仕組みで誰がつけていくのか。(構成員)

▶ 今回のように企業が設置する考えに至った背景としては、現状のセンサーの取扱いは役所の中でもバラバラであり、例えば防犯カメラは補助金で自治会が設置し、鳥獣被害対応用は別のセクションが所管するなど、機器の統一やデータの共有は行われていないことがある。企業側がビジネス化を目指していることもあり、これらを解決してくれるのは企業のアイデアや技術ではないかと考えている。(神戸市)

- 統計局ではオンサイト施設などで統計のマイクロデータを提供しており、例えば自治体が行政記録情報を持ち込んで、統計の個票データ、マイクロデータと一緒に分析することができる仕組みとなっている。一定の教師データがあることが前提ではあるが、行政記録情報に将来の空き家確率や成人病罹患確率などの確率変数を付与することが可能である。現時点では、マイクロデータ自身を持ち出せるのは、統計目的利用に限られているが、仮にこれを統計目的だけでなく、行政目的で持ち出しが可能になれば、有効活用できると考えるか。（総務省）
- 個人情報保護の観点などから、大学に研究目的として公共データを出すのはなかなか難しいだろうと考えている。出すためには、大学と連携協定を結んで共同研究とすることになるが、研究目的と行政目的の両方を満たすテーマにし、都市計画の分野など、限定的な形になるが、できるだけ出す方向で考えていきたい。ただし、クリアしないとイケない様々なハードルがあると思う。（神戸市）

以上