

## 第9回 統計委員会委員と統計利用者との意見交換会 概要

1 日 時 平成27年4月23日(木) 10:00~11:05

2 場 所 中央合同庁舎第4号館12階 共用第1208特別会議室

3 出席者

### 【委員】

西村委員長、北村委員長代理、川崎委員、西郷委員、白波瀬委員、中村委員、中山委員、野呂委員、廣松委員、前田委員

### 【統計利用者】

大口 敬 東京大学生産技術研究所教授

### 【国または地方公共団体の統計主管部課の長等】

内閣府経済社会総合研究所長、内閣府経済社会総合研究所総括政策研究官、総務省統計局長、総務省統計局統計調査部長、財務省大臣官房総合政策課調査統計官、文部科学省生涯学習政策局政策課教育分析官、厚生労働省大臣官房統計情報部長、農林水産省大臣官房統計部長、経済産業省大臣官房調査統計グループ調査統計審議官、国土交通省総合政策局情報政策課長、日本銀行調査統計局参事役、東京都総務局統計部長

### 【事務局等】

伊藤内閣府大臣官房統計委員会担当室長、佐藤内閣府大臣官房統計委員会担当室参事官、田家総務省政策統括官(統計基準担当)、小森総務省政策統括官付統計企画管理官

4 議 事

(1) 統計委員会担当室からの報告

(2) 統計利用者からのプレゼンテーション

大口 敬 東京大学生産技術研究所教授

「交通部門におけるビッグデータの整備状況と課題」

(3) 意見交換

5 資 料

資料1 事務局資料「第9回統計委員会委員と統計利用者との意見交換会について」

資料2 大口 敬 東京大学生産技術研究所教授

「交通部門におけるビッグデータの整備状況と課題」

6 議事概要

(1) 統計利用者からのプレゼンテーション

- ・本日は、交通分野における行政所掌とデータの種類の状況、データ・ソースにはどのようなものがあり、ICTによってどのように変わってきているかということ、さらに課題と展望の3点について説明したい。

#### (行政所掌とデータの種類)

- ・交通は大きく陸上、海上、航空に分けることができる。陸上については、道路を作っている国土交通省道路局が道路法を所掌、交通ルールを作っている警察庁が道路交通法を所掌している。また、道路輸送に関しては、同省自動車局が道路運送法を所掌していて、同省総合政策局は、全体的に司っている。同じ陸上でも鉄道は鉄道事業法、路面電車は軌道法と分かれている。このように交通に関する行政官庁はいろいろあり、環境影響に関しての排気ガス規制や騒音規制になると環境省も関わってくる。
- ・統計情報については、自動車交通の分野においては、道路局により日本全国で一定のルールの下、道路交通センサス(全国道路・街路交通情勢調査)が行われ幅広く交通の実態が把握されており、自動車局により自動車輸送統計調査が行われ、バスやトラックの台数や用途を把握している。貨物の移動に関しては、総合政策局で物流センサスを行っている。一方で、人の動きになると、都市圏では都市圏パーソントリップ、地方都市では全国都市交通特性調査(全国パーソントリップ)が行われている。交通事故、免許の保有については警察庁、騒音、環境影響になると環境省が所掌している。鉄道分野は、大都市交通センサスで調べている。このように所掌に応じて、統計もいろいろな部署に分かれている。

#### (データ・ソースとICT)

- ・自動車交通では、トラフィックカウンター(略してトラカン、車両感知器、検知器のこと)を使ってデータを取っている。路面にループコイルを埋める、あるいは超音波センサを置いて車の存在を検出している。トラカン通過で交通量を、またトラカン2つを極めて隣接して置くことで、その2つの通過のタイミングのずれから速度を測っている。日本は戦後にこのような機器が大量に整備されたので、相当なデータが取れる。交通信号も同じような原理に基づき、車両感知器を使い交通管制センターがたくさんの信号を集中制御している。最近ではプローブという車の移動をトレースしていくような技術、データを活用して、渋滞に応じて、全体の信号を調整するなどの制御をしている。車のプローブ情報によって車の移動をデータ化することができ、GPSや様々な加速度センサの情報と組合せると、急ブレーキが多い地点やヒヤリハットが多い地点などから事故が起きそうな地点などを察知することに利用できる。ETCは高速道路利用車の9割まで普及していて、ゲートオン、ゲートオフのタイミングによって旅行時間が分かるし、どこからどこまで行く車がどれくらい多いかということ把握するのにも使われている。しかし、これらのデータは、交通の統計作成、信号制御、あるいは料金収集のためという目的で収集されているので、生の情報のままではなく集約化されており利活用上有益な情報が削ぎ落とされている状況にある。年間の交通情勢の把握や記録保持のためだとリアルタイムの活用は難しく、逆に信号制御のためだとリアルタイムで活用する目的なので必ずしもデータを保存していないという状況にある。
- ・最近、ビッグデータという観点から、画像をデータ化することが増えており、融合化されれば交通流の全体像を描くことが可能になるだろうと思う。リアルタイムな活用のニーズはいろいろなところであって、交通が集まって混んでいるところを分散してもらうように瞬時に何らかの指示をすることが可能になると思う。これを実現するには、情報の蓄積とその解析から得られた傾向・特徴が抽出され、これとリアルタイムで収集されるデータがうまく融合して活用される必要がある。

- ・トラカンという従来型の感知器は、断面の通過の台数、地点の速さは分かるが、道路上にどれだけ車がいるのかという車の密度は分からない。最近、インフラコストの縮減から、センサの設置台数は減少傾向にある。カメラさえ設置しておけば画像でどうにかなるかもしれないが、画像処理は車が陰に入ると台数計測にエラーが生じやすい。一方、移動体タイプのプローブは、車メーカーなど民間主導の動きで普及しており、車の時々刻々の動きが分かり、インフラ投資は不要という特性があるが、サンプル調査となるので通過台数を知ることは難しい。公的セクターが活用する際には、通信データ量が膨大なので通信料金の問題、緯度経度から道路を特定することをマップマッチングというが、それをしないと利用できないという問題がある。また、1km単位の平均的な速度にするのか、1秒単位の計測された生のデータにするのか、集計方法も一定のものはない。さらに、市場競争が成立していない制度設計であり、データ保有者の言い値で買わざるを得ないのではないかと個人的には思っている。
- ・ICT化が進んでハードディスク等が安くなってきているので、データの蓄積、詳細な集計が可能になっている。蓄積したデータから平均傾向を把握したり一般化したりすることもできるので、汎用的な利用が可能である。いろいろな異なったデータ・ソース間で融合的に活用されていくのが大きな方向性ではないか。例えば、個別のセンサはそれぞれ良い点と悪い点があるが、それらを組み合わせて、いわゆるデータ同化のようなシミュレーション技術へと進展する方向にあると思う。交通というのは、人や物を動かす場合でも社会的な活動等に大きく影響を受けるので、一見すると、交通とは直接関係のない天候やイベントのようなものとリンクできると思う。最近でいうと、ツイッター等のSNSから交通渋滞の状況が分かるというような研究もある。
- ・個人の動きをトレースする形のパーソントリップは、アンケート調査で自己の行動を思い出して記録するというもので、観光なのかビジネスなのかという目的、あるいはなぜ車に乗らずに電車で行ったのかという手段選択の理由などを聞くことができるが、数年に1回1日限定で全国一斉にやるパーソントリップの場合は、交通の全体像把握にはサンプルが極めて限定的である。そこで現在、ICTを使って調査する話がでている。パーソン・プローブというが、携帯やGPS、Bluetooth/wifiなどさまざまな技術を用いて個人の動きを把握する方法である。交通系ICカードを使えば、どこからどこまで移動したか分かるしサンプルも増やすことができ、毎日のようにデータを取ることもでき、捕捉できるトリップも増えて経路もある程度分かってくる。ただし、ICカードを使っていれば別だが、GPSを見ている限りでは、自分で車に乗っているのか、バスやタクシーに乗っているのか区別するのは難しい。したがって、何らかの意味づけのデータを取らなければならないが、そう簡単ではない。例えば携帯の場合では、データを通信キャリアから入手する必要があり、プライバシー問題もあってこれは極めて難しい。
- ・運送のうち旅客事業については、安全性等を確保する観点もあって動態を把握できる仕組みがほぼ整っているといえると思う。バスの場合、バスロケーションシステムでバスの現在地情報をサービスとして利用客に提供しているが、これはバスの管理のためにも使われている。バス車内の乗車率は、ICカードの導入により把握できるはずと言われているが、東京のようにほとんど料金が均一な区間だと降車情報がないため、結局、調査員が同乗して乗車人数を調査しないと分からない。貨物事業も同じで、事業者の管理システムとして移動状況はデータ化されていると思うが、どんな荷物をどのくらい、どこに移動させているのかは分からない。荷物は、ICタグでこの配送センターからこのお店に運ばれたのかフォローできるとは思うが、それを集約してこのトラックに何が載っているかをみることは容易ではない。トラックの重さは、料金所で車重計

により把握しているが、過積載の検出に使うのみで、過積載ではない場合はデータ化されていないと聞いている。

#### (課題と展望)

- ・東京周辺で3環状道路と呼ばれる高速道路ネットワークが実現しても、どこで渋滞が起きるかということは体系的に知識化されておらず、交通容量がどれくらいかもよく分かっていない。交通が超過しそうなら他に振り分けなければいけないが、需要を知るということは非常に難しい。近未来の状態を予測できれば、道路を有効活用することが可能になると思う。そのための手法論は、経路誘導、需要調整、動的交通管理、インシデント対応である。
- ・渋滞対策、事故対策の研究を行うためにはデータ収集と分析をしないといけないが、そこに時間と費用が相当かかっている、全体の7、8割を費やしているというイメージである。ビッグデータ時代の観点からすると、必要な情報と標準化された分析手法があらかじめデータベース化されて、課題抽出や対策検討がより高度化される方向にあると思っている。
- ・例えば交通事故について、どこで事故があったか、事故にひも付けられるいろいろな要素として性別、年齢、ぶつかり方、車の種類等、事故が起きたときの天候、路面や渋滞状況などの要素を全部データベース化すると、ある条件で事故になってしまった、それはどのような事故であったかということが分析でき、その場所は事故多発地点といえるかどうかということが分かるようになると思う。
- ・公的資金が減っていく中で、民間をうまく活用する観点が大事だと思うが、交通に関する多くの民間データは市場原理に基づいた価格設定がされていない。東日本大震災のときは、民間企業の枠を超えて各社が保有するプローブ情報を集約することによって、あの道路に車が1台も通過しないのは、あの道路が壊れているからに違いないというような使い方がされた。このようなデータ活用は、交通の状態を把握する公的なソフトインフラの一つといってもよい。ただし、このようなデータは品質とセキュリティの問題がある。

#### (2) 事務局からの調査研究のまとめに関する説明

事務局から、資料1に基づき調査研究について説明が行われた。

- ・交通部門においては、車両感知器やセンサなどを通して官と民とが集めている多くのビッグデータがある。これらを利用することで、経済動向を把握できる可能性は一定程度あるものと考ええる。
- ・ただし、課題もある。例えば、有料道路に関しては、複数の道路会社においてデータ様式が統一されているのか、車種別として大型と中型の区別は可能だがトラックかバスかを区別できるのか、さらには割引料金制度が種々あるので料金収入から輸送量を把握するには何らかの工夫が必要という問題などがある。データを利用する場合、匿名化された個々のデータをもらって必要な加工をするのか、あるいは加工処理後のデータをもらって利用するのかという問題もあり、どちらもコスト負担の問題が発生する。
- ・宅配業者の取扱量というデータに通販業者情報などの既存のデータを突き合わせるような利用方法も考えられる。
- ・ビッグデータを活用することは、既存の統計よりもコストが安くなる、速報性が高くなるというメリットがある。ただし、ビッグデータは民間、公的機関を含めて分散して保有されており、

それらのデータの整合性という問題がある。

- ・ 公的機関としては、ビッグデータは役立つ、こんな分析ができるということを情報発信していくことも重要な役割である。

### (3) 意見交換

- ・ 研究で民間のデータにアクセスする場合に、研究者だから特別にというようなアレンジができるものなのか。

→今の段階では、民間会社もどうやればデータが価値を生むのか探りたい状況のようだ。交通に限らず、気象や農業を研究している方のところに民間の方がアクセスしてきて、自社のデータの価値を高める方法について、共同研究の依頼をすることがある。ただし、企業側は、個人情報保護に気を遣っていて、私どもが民間会社にデータ提供依頼をしても、多くは難しいということになるのが実態だと思う。

- ・ 今回ご紹介いただいたデータの中に、準公的な機関が持っているデータもかなりあると思うが、そういう場合は制約があるものなのか、それとも、頼めば利用させてもらえるものなのか。

→公的データに関しては、所掌の行政行為の中での責任の範囲内でデータを持っているということなので、これを違った使い方をされることに対して、非常に防衛的に考える官庁も結構多いと思う。そういう意味ではやはり一定の歯止めがかかっているのが実態だと思う。

- ・ 公的機関といってもいろいろ分散しているようだが、データをリンクして使う場合、例えば道路でも所管が違うので、支障はあるのか。

→最近、特に中央政府は省庁をまたいで目的を達成するための研究プロジェクトも実施しており、その中では異なるデータをリンクさせて活用するという話が進みつつあると思う。

- ・ 民間提供データが有償購入になっていて、市場原理が働いていないので、結構高い値段をチャージされるという話があったが、一つは供給側が限られていて、独占的に情報を集めていて、それ以外の参加者がいないということだと思うが、それは技術的にその特定の業者が優れていて、情報を集めているという状況なのか、それとも参入障壁のようなものがあって、競争者が入れないのか、そういうものがビジネスになるということがあまり認知されていないのか。

→私の認識は、主には最後の3点目ではないかと思う。例えば、自動車メーカーは、顧客サービスの目的でプローブから車が時々刻々どこを走っているかというデータを取っているが、そのデータがあると旅行時間、渋滞の場所などが分かるため、交通調査したい機関があるメーカーにデータ提供を依頼した。その結果、買ってくれる相手がいるなら商売になるということでそのメーカーだけがデータを加工して提供するようになったというのが実態だと思う。

- ・ プローブの情報は、消費者が買った車の中に装置が付いていて、自分の行動が逐一見られているということを消費者が了解した上で、データが取られているのか。今後、公的統計で利用するようなことを考えた場合、データを提供する側が意識して売買契約書のどこかにそういうことを書いておいて、消費者がサインしたことをもって認めたという話になっているのかもしれないが、消費者情報の保護という観点から、会社側はどう考えているのか。

→正にプライバシーの問題ということで、非常にセンシティブであり、個別車両のデータは一切表には出てこない。したがって、完全に個人が特定できない形においては、活用することもあるということをユーザーとは契約しているはずである。メーカーは、ユーザーから取得した大量のデータを集約して、混雑地点をカーナビに表示してユーザーに返すというサービスをする

ためにデータをとっているの、ユーザーに返している形のデータを集約したものであれば、一定のルールの下で第三者に提供しても良いだろうというような考え方だと思う。

- ・技術的なことだが、自動車輸送統計で距離のデータというのは結構正確に取れる。運転している人に調査票を配っても、メーターを見て書いてくださいと言っているの、記憶に基づいて書いた数字よりは正確なものが取れる。それに対して、同じ自動車輸送統計の中で、ガソリンの消費というのは、環境の観点からかなり高い精度が求められるが、精度の面から不安が残るような調査方法である。エネルギーの消費量の把握という観点から、データが使えるか。

→車は、電子制御されるようになったことで、実はいろいろなデータを内在している。車の中にLANに相当するCANというものがあり、エンジンが何回転回り、どこでどれだけの燃料を噴射したというデータも作られていて、それらを蓄積しておいて整備のときに活用している。車が持っているデータがビッグになっているので、日本全体でトラックの輸送に使われているエネルギーの消費がどうなっているかということに活用できる可能性がある。しかしながら、一回転ごとの噴射量を膨大に取ってもそれだけでは意味が分からない。例えば1日に噴射した総量がほしいのだが、それは簡単ではない。燃料の残量センサをある一定の方法でモニターしていくことで、どれだけ消費されているのか分かるはずであるが、それを累積しようと思うと、一定の方法で保存するか、あるいは加工したものを保存するなど、新たな問題が生じる。それができないと、多くのデータが無散していくことになる。

- ・パーソントリップについては、社会学という観点からいうと、非常に読みにくいという感じを最初に受けたが、今日お話を聞いていて、キーはどう組み合わせるかということだと思う。面と点をどう組み合わせ、自分の見たい目的を可視化していくかということが最終的に一番重要と思ったが、その時にそれぞれの官庁が違くとデータの定義や調査目的も違うので、結局データの単位も違ってくる。これが民間になると統計としての基本単位自体が非常に多様化してしまう問題があると思うが、最終的に結果を発表する研究者としての責任でよいのか、データの管理責任自体を分散化しながら、ビッグデータを使っていくという方向性がよいのか。

→データベースという概念をどうイメージするかということとリンクしていると思う。データベースという言葉は初めて聞いたときは、データを集めてくることだというイメージがあったと思うが、今は、データ自体は分散的に持たれていて、それを上手につないで、そこで意味を読み出すような仕組みをもっていればそれがデータベースだと解釈すべきと思う。そういう意味では、分散的に持ち、データの質を含めたメタ情報も管理していただいて、それを一定のルールの下で、融通していただくようなデータベースの持ち方を工夫していくことが重要だと最近感じている。

- ・携帯電話の場合、事業者側が集めたデータを単に事業の効率化、あるいは運用するために用いるだけではなくて、官側に利用させるとか、民と民の間で融通して利用するとき問題になったのは、データをとられた側の意識への対応が難しいということだった。交通分野の仕事で、クレームあるいは問題提起を直接経験したことはあるか。

→私は直接関わらなかったが、Suicaデータの漏えい報道があったことによって、鉄道のICカードの利用はほとんどできていない。これは社会的な制裁が生じてしまったという状況だと思う。一方ではAmazon、Googleを使っていると便利だが、自分のデータを取られているという気持ち悪い状態にある。自分のデータを適切に使ってもらうことを認める社会制度やルール作り、あるいは個人の啓発などに総合的に取り組むことが必要なのではないかと感じている。

- ・データを組み合わせる場合のシミュレーションというのは、どのようなものがあるのか。
- 交通の分野のシミュレーションは、かなり多様なシミュレーションを含んでおり、例えば20年後の交通予測をするための交通需要予測のシミュレーション的な計算というのあれば、時々刻々とまるで交通管制センターで交通を見ているかのようなことを計算機でリアルタイムに変動をシミュレートする、あるいは合流部での車の挙動をシミュレートするというようなこともあり、いろいろな要素のシミュレーションをやっている。

以上

<文責：内閣府大臣官房統計委員会担当室>