

国土交通政策研究所 第170回政策課題勉強会 概要

日時：平成26年12月17日(水)12時30分～14時00分(※実際は15分程延長)

講師：国際大学 グローバル・コミュニケーションセンター(GLOCOM)

准教授・主任研究員 中西 崇文 氏

講師・主任研究員 庄司 昌彦 氏

テーマ：ビッグデータ・オープンデータの活用の現状と国土交通分野

1. 中西氏「ビッグデータ活用概要編」

(1) ビッグデータの定義と現状

- 最も一般的な定義は「3V」。即ち、「Volume」でデータの量が大きいこと、「Velocity」でデータの更新頻度が高いこと、「Variety」でデータの種類が様々であること(テキスト、音声、センサー、画像等)の3つを充たすデータを指す。
- 感覚的に述べると、現在1秒間にインターネット上を流れているデータ量は、1990年台のインターネット全体のデータ量と等しい。また、人間の脳の容量は3TBと言われる一方、現在のICT上で1日に生成されるデータ量は2.5EB(エクサバイト)=2,500,000TBであり、まさに桁が違う。
 - Cf.) 8つのON/OFFスイッチがあれば、英数字と記号を表すことができる。ON/OFFスイッチ1つを示す単位が1bitであるため、1B=8bitが情報の基礎単位とされた。
 - Cf.) 1日のデータ生成量は、Facebookが「いいね」27億件・写真3億件、Twitterが投稿4億件、Googleが検索10億件、楽天が注文62万件、Amazonが注文3,600万件、LINEがメッセージ70億件。
- ICT機器に記録されている情報量は2013年・4.4ZB(ゼタバイト) → 2020年44ZBに、ビッグデータ市場は2013年・148億7,000ドル → 2018年・463億4,000億ドルと拡大すると見込まれている。ユーザ企業がビッグデータに関心を持ち始めている今、アナリティクスニーズの吸い上げを図るべきと言われている。即ち、データの利活用を考え、ビジネスに結びつけるチャンスを創るべきだ。

(2) ビッグデータの重要性

- ビッグデータが重要になった理由は2つ。
 - 第一の理由は、コンテンツの作り手が変わったこと。従来の作り手は、作家、漫画家、マスコミ等の「特別な者」であった。現在は、UGM(User Generated Media)の登場によりユーザー全員(=一般の人)が作り手となることができるようになり、更に、Social Mediaでユーザーが繋がることでコンテンツの流通が激化した。
 - 第二の理由は、センサーの廉価化であらゆる場所にセンサーが置かれる様になったこと。例えば、携帯電話の中にもGPSセンサー、輝度センサー、加速度センサー等様々な

センサーが搭載されている。これによって、現実のより詳細な写像が可能となった。

- ユーザーが情報を発信するということは革新的で、ユーザーレビュー等を通じて、企業とユーザーとの情報の非対称性が回避されるようになった。また、お互いがお互いよりも詳しい分野があることにより情報の多様性・情報の格差が生まれ、情報が流通し続けるようになった。

- 資料15頁緑色の線の通り、UGMを活用する顧客(ユーザー)は増えている。つまり、従来の消費体系とは異なり、企業が発信する情報(ブランディング)ではなく、消費者が発信する情報(ロコミ)が重視されてきている。

- ビッグデータを巡っては波が存在する。

第一の波はインフラストラクチャの整備だが、これが無償公開されたことがビッグデータの流行に火を付けた。この段階は、既に通り過ぎている。

第二の波はセンサーデータ生成で、データ量の増加だ。これで様々なデータを取得することが可能となり、データを処理できるようになった。

現在は第三の波の段階にあり、「何を処理すれば良いのかわからない」という時代だ。データの利活用により付加価値創造を模索することが求められる。

将来、第四の波としてプラットフォーム、即ちビッグデータを交換し合う環境を誰が構築するかが問われる時代が来ると考えている。どの情報を交換、シェア、オープン、クローズするかという選択が重要である。プラットフォームが構築されれば、社会が変わると言われている。

(3) ビッグデータの活用・分析

- 従来の分析は、「仮説検証型」といわれるもので、仮説を立案して証明するというプロセスを踏む。その際、母集団の推定やサンプリングといった手法が用いられる。しかし、ビッグデータ分析は全く異なり、仮説が存在しない。母集団≒データであり、データの世界が現実の写像と考え、そのまま分析する。データが現実の写像であるとするもので、数学でいう「大数の法則」で認められている。

Cf.) 従来の分析で仮説を立案していたの背景には、センサーが高額であったため、設置できる箇所が限られていたという制約があった。そのため仮説を証明できるようにセンサーを配置していた。

- 数千年前の経験科学、数百年前の理論科学と実験科学、数十年前の計算科学を経て、ビッグデータが属する「データ中心科学」が生まれた。計算科学の一種であるシミュレーションは、ビッグデータとは異なる。データ中心科学にモデルは無い。モデルがないなかで、現実がどうなっているのかを分析する。

- Who、How、What、Which、Whereで整理するのがビッグデータの分析手法となる。データは現実の写像であるのだから、現実同様に整理する。この際、データの取得や整理をし続けることがポイントとなる。つまり、現実を把握し続けること、顧客のニーズを把握し続けることである。そして、そのためには現実世界からサイバー世界へデータを送り続け

るインターフェースが求められる。

Cf.) AppleのSiriがビッグデータを生み出すインターフェースを抑えたことによって、キーボードやマウスを使えない若者が増加した。インターフェースから価値が生まれ、未来が変わる。

- 紙面上へのビジュアライゼーションに当たっては、2軸又は3軸が限度となるため、どのような軸を選ぶかということになる。(グラフ種類の選択の作法については資料27頁参照。)
- データ利活用は情報システム部門だけの仕事ではなく、全員が考える必要がある。スケールメリット(規模が大きい⇨現実の写像)、スコープメリット(多角的な視点)、コネクションメリット(データの繋がり)を基に考えることになる。

Cf.) スケールメリット … GEのガスタービン

Cf.) スコープメリット … JR東日本ウォータービジネスのacure

→ 「温度」を「消費者心理」と置き換えて商品入替

Cf.) コネクションメリット … 大日本印刷によるレシートとパーソナルデータの結合

- 顧客像の捉え方として、エスノグラフィ、プロフィール、ペルソナという3つの言葉がある。多量のセンサーを設置することでエスノグラフィを明らかにすることが可能となるし、近年はビッグデータからペルソナを明らかにして欲しいという企業が多い。
- 情報コストが下落すると情報速度が上がる。ハッブルの法則(遠くの天体ほど速いスピードで遠ざかる)はビジネスにも当てはまると言われており、企業からの距離が遠い顧客ほど速いスピードで企業から遠ざかると言われている。ここでビッグデータを活用しなければ、ビジネスに負けてしまう。

(4) ビッグデータによる価値創造

- 意味があるものに価値があるとすると、意味を見出すことが価値を創ることとなる。
- 「意味」はコンテンツとコンテキストの2つから成る。コンテンツとは明示的に表現された内容、コンテキストは背景的に表現された文脈。ビジネスに置き換えると、コンテンツが商品、コンテキストが背景や場となる。
- 従来は、モノを消費するということが豊かさの象徴とされ、コンテキスト(背景、場)が共通していたため、コンテンツを伸ばせば売れた。しかし、現在は、技術発展が緩やかとなり価格競争によってコンテンツの差別化が困難となった。消費者の多様化によってコンテキストも共通ではなくなったため、今の時代に価値を創造するにはコンテキストが重要である。企業はユーザのコンテキストをビッグデータによって(ニーズとして)捉えようとしているが、これは消費社会からの脱却を意味する。つまり、ビッグデータは消費社会からの脱却を意味することになる。
- Dirtyな(構造化されていない)データを対象として分析できる点がビッグデータの技術的コアである。ビッグデータの技術的・工学的アプローチはオンライン学習(頻繁に入力されるデータを分析しながら人工知能の学習にも利用する)と特徴選択(よりよい結果を出す特徴量とは何かを抽出する)の二つしかない。従来の分析技術では、オンライン学習(学習に時間がかかるため、頻繁な環境変化に対応できない)や特徴選択(機械は、眼等の特徴から人間の顔であると認識するこ

とができないが困難だったが、現在、これを自動的に行う分析技術が注目されている。

(5) サービスとプライバシー

- プライバシーを守るため、1)匿名化、2)告知と承認、3)オプトイン／オプトアウト、という3つの方策が採られる。しかし、これらの方策での対応には**限界がある**。

Cf.) 匿名化をしても、データを統合すれば特定可能である。(単に時間がかかるだけ。)

Cf.) その都度その都度告知と承認をすることは非現実的・非効率的である。(交通系ICカードの目的変更の度に改札口で止めて承認を求める訳にはいかない。)

Cf.) オプトイン／オプトアウトをすとかえって目立ち逆効果となる。また、「忘れられる権利」として検索サイト上の処理をしても、元データの削除まではできない。

- 東日本大震災の際に「被害に遭った避難者の多くは海の方へ避難していた」という報道があった。違和感の無い報道に聞こえるが、この結論を分析によって導こうとすると、点のデータであるGPSでは「海の方へ避難」というレベルまでは不可能だ。つまり、GPSのデータと一個人を特定する情報とを紐付けた上で、時間的に連続させて追跡する分析が行われたということになる。

日常生活、例えばどこへ食事に行った等のレベルで同じ事をされたらと考えると非常に恐ろしいが、この報道(分析)については社会的に有意義なものとされた。

- 提供するものと取得するもの、サービスとプライバシーのバランスが非常に重要となってくる。
- Suicaの件では、サービスとプライバシーのバランスという視点に問題があったと言える。

(6) その他 / 全般

- 「パーソナルデータ」は「個人情報」とは概念が異なる点には注意が必要。概念図は資料48頁参照。
- 「ビッグデータマガジン」にて、1ヶ月に1度程度の頻度で執筆をしている。

2. 庄司氏「オープンデータ活用編」

(1) オープンデータの定義と現状

- ポイントは自由に使えること。細分化すると、1)オープンなライセンス、2)オープンなアクセス、3)オープンな形式、の3点となる。

Cf.) 「オープンな形式」とは機械が使いやすいこと。スキャンしたデータよりはcsv形式のデータ、有償ソフト無しには扱えないデータよりは無償ソフトで利用できるデータがオープンでなる。

- よく誤解される点だが、単にウェブ上で公開することを指すものではない。自由な利用というだけではないし、数値だけではなくその他データも含むことが必要である。台湾の表現を用いるなら「公表」ではなく「開放」である。

- オープンデータの議論の対象には、個人情報や機密情報は含まれない。既に公開されている情報又は公開可能な情報の中で、自由に使えるものが少ない、増やすべきである、もっと使いやすい形で発信すべきである、という議論である。
- ビッグデータは大きさの問題、オープンデータは利用可否の問題であり、全く次元の異なる話である。結果として、両者は概念的に重なり合う。
- また、行政との関連では基本的に、オープンデータの活用はGtoCとGtoBの形での活用となるが、ビッグデータはinG(政府機関内)やBtoGが主に議論されている。
Cf.) BtoGの例として、ホンダのデータが埼玉県交通安全に活用された例が挙げられる。
- 行政が全てに対応することが難しくなってきた現在にあつては、行政が保有する資源を提供し、民間による自律的課題解決を促すために、オープンデータというものが役に立つ。
- ソーシャルメディアの登場とIT環境の低価格化により、NPO法人等の小さな主体が機動的に活動する環境が整ってきている。オープンデータは彼らの活動資源になる。
Cf.) 動画の撮影・放映にしても、以前はマスコミにしかできなかったが、現在は携帯電話1つで個人が行うことができる。

(2) オープンデータの国際的動向

- 米国が注目されがちではあるが、英国の立ち回りが巧みである。(資料14頁参照) ブラウン政権からキャメロン政権に政権が交代したが、オープンデータの方針は引き継がれている。
- G8オープンデータ憲章の技術的別添では、提供を進めるべき情報のリストが示されており、これに基づいて日本政府もアクションプランを作成した。8ヶ国が同じデータを出していくことで、国を越えての比較もできることになる。
- オープン・ガバメント・パートナーシップが英国の熱意の下に進められ、64か国が参加しているが、加盟後に脱退したロシアを例外とすれば、G8ではドイツとともに日本が参加していないという状況にある。
- 民間の取組であるオープン・ナレッジ・フェスティバルは、これまで8回開催されている。そのうち、2012年の会合では透明性の話題に終始しており、成長戦略に位置づけていた日本の考えとは少し異なるものだった。2013年の会合には発展途上国が多数参加し、英国と世界銀行の主導で発展途上国への支援が決められた。2014年の会合では、2013年の会合の支援の成果が発表されるとともに、さらに多くの発展途上国が参加し、この段階で経済が議論された。併せて、G20にG8オープンデータ憲章への署名が要求された。

オープンデータの話は、ともすれば、トイレに関するアプリの開発等、一地方の話と認識されがちであるが、それが世界と繋がっていることを忘れてはならない。

- 日本政府の取組は、改善されていると評価されてはいるものの、高い評価を受けている

訳ではない。取り組むべきとされている点は明らかで、1)公共交通機関の時刻表、2)政府支出のトランザクションレベル(より細かいレベル)での可視化、3)全体的なオープンライセンシ化の3点である。

Cf.) 日本では公共交通機関の運営主体は民間であるが、国際的な議論では、公共交通機関の時刻表は公共データとして扱われる。

(3) オープンデータに係る国内での政策展開と民間の取組

- 2011年東日本大震災が大きな契機であり、2014年と2015年の2年間は集中取組期間と位置づけられている。

Cf.) 2011年東日本大震災が契機となった取組として、電力の需給状態データがオープンとなり、様々なアプリが自発的に開発された。通行可能箇所・不能箇所や緊急地震速報といった優れたものも挙げられるが、残念ながらこれらはオープンデータとはなっていない。

- 国・地方公共団体の取組は資料27頁～35頁参照。特に殆どの政令指定都市は何かに着手しており、アンケートの結果以上に熱心である。
- 世界同時に開催したInternational Open Data Dayでは、世界194都市のうち32都市が日本であり、コミュニティ単位の活動は日本は盛んになってきている。年明け2015年は2月21日(土)に開催予定で、現時点で27、28都市での開催が予定されており、昨年よりも多数となる見込である。
- Code For X(地域名)も盛んで、既に繋がっている地域は18地域、途上にある地域を加えれば約30地域となっている。

(4) 事例 ～ 国土交通分野を中心に

- 米国の「MRIS」は、不動産に関するありとあらゆる情報が非常に詳細に掲載されているもの。元となっている情報は全て公共機関のものであり、それを巧く組み合わせている。
- ビジネス的な成功事例として必ず挙げられるのが、米国の「Total Weather Insurance」。気象、過去の収穫量、土壌のデータを掛け合わせ、農作物栽培に関するリスクを自動的に算出し、保険を掛けることで、農家の収入の安定化を図っている。元Google社員が設立し、多国籍バイオ化学分野の巨大企業Monsanto社に11億ドル超で買収された。
- 千葉市の「ちばレポ」は、課題を抱え込まずに市民と共有し、ゆくゆくは市民による自律的解決に繋げようというものである。
- 郵便事業株式会社の子会社であるJPメディアダイレクトの「カスタムタウンプラス」は、従来の住所指定による全戸広告配布を、住民属性等の様々なデータと組み合わせて広告の効果を上昇させるもの。同種のサービスは、大日本印刷等も手掛けている。
- 株式会社おたにの「GEEO」はMashup Awardsのオープンデータ部門賞を受賞した。多様な空間情報から、築年数等も加味した不動産価値を算出する。

- 明石工業高等専門学校が作成した「NGY Night Street Advisor」は経済産業省・総務省のコンテストで最優秀賞を受賞したもの。名古屋の街路灯の配置や電球の種類から夜道の明るさを推定し、Google Mapの上に重ねて安全なルートを検索する。
仮に、Google社がこのアプリを買収すると、Google Mapのルート検索に「安全モード」というものが備わるかもしれない。
- 「台風リアルタイム・ウォッチャー」は、ウェザーニュースのロコミ情報と国立情報学研究所の情報を合わせて視覚化している。元々の対象は台風だけだったが、ありとあらゆる災害に有効ということで幅広く活用されている。
- その他、税金の使途を可視化する「Open Spending」は非常に関心が高い。特に日本からのデータ投入が多く、日本では草の根的な活動が盛んな様だ。福岡市の統計サイトではインフォグラフィックによる市のプロモーションと統計データ提供を行っている。金沢市の「5374.jp」は、市に多く寄せられるゴミ関係の問い合わせへの対応を行ったもので、シンプルではあるが人気があり、50都市程に広がっている。

(5) 今後求められること

- まずはデータを出すこと。出せないデータを出すということではなく、出せるデータを出す。それにより、民間の創意工夫が後押しされ、様々なサービスが創出される。(資料52頁参照)
- ニューヨーク大学のデータによれば、オープンデータを活用している分野の上位に「Finance & Investment」、「Business & Legal Services」、「Governance」といったものが挙げられている。所謂「技術の企業」ではなく、アルゴリズムに価値を見出している者が多い。IT企業とは異なる者と相談していく必要もある。
- 「Total Weather Insurance」の様に、複数のデータを掛け合わせるということが、成功事例の中からは見て取れる。それによって新しいものを生み出せるかどうかが鍵となる。
- 社会的インパクトが大きいのは、GtoBtoBtoC、即ち、ビジネスを支援するビジネスだ。
Cf.) もしかしたら既存組織にとっては「中抜き」ということになるのかもしれない。
Cf.) 有名な徳島県上勝町の事例も、どの葉をどれだけ出荷すると高く売れるかというビジネスで、ビジネスを支援するビジネスと見ることができる。道の駅の事例も同様。
- 情報公開請求の多くは企業からの請求であり、その内容は金額入り工事設計書。結果的に殆どが開示されており、そうであれば最初から提供して良い筈
- 課題としては、著作権とプライバシーが二大課題と言える。

(6) その他 / 全般

- ビッグデータと言うと、ともすれば「データサイエンティストが扱うもの」と他人事に捉えられがちである。しかし、“スモールデータ”、即ち、個人のパソコンに保存されており、Excelデータ、Word文書、写真といったデータが自由に、多くの人に使われることの方が革命的と言われている。
- 現在行われている個々の取組は小さなものに過ぎない。しかし、イノベーションを起こす時には多くの小さな試行錯誤が無ければならず、それを育てる環境が必要となる。

3. 質疑応答

[問1]

「ちばレポ」は、オープンデータではあるが、市民から情報を提供してもらっているという側面もある。財源が限られ、報酬を払うということも難しい中、市民や企業から情報を提供してもらうための知恵があれば伺いたい。

[答1]

その点は大きな課題である。「ウェブによる参加型であれば人が寄ってくる」などという簡単な話ではない。

とは言え、自発的にやる気のある方々に受け皿を提供したということは、意義の1つとして挙げられる。その他としては、千葉市では「ちばレポ教室」というものが開催されており、人が集まって勉強するということがモチベーションの1つとなる。市長が参加しての街歩きも、実際の発見が生まれるという点でモチベーションを生み出している。(高齢者の場合には、スマートフォンで写真を撮ってアップするという体験をこれを機に学んでみようということもモチベーションになっている。)

その他、街路灯に番号を振って特定の番号を探すゲームを開催したり、サッカーチームとタイアップした企画を開催したりということも為されている。

[問2]

Suicaの事例は大きな話題となったが、不特定多数とのオーソライズはどの様に進めていけば良いのか？

[答2]

Suicaの事例では、個人情報保有するJR東日本が、匿名化した情報を日立へ提供した。個人情報保護法の観点から言えば、照合容易性の点で抵触するものである。この照合容易性は情報を渡す時点が判断基準となる。匿名化された情報を受け取った日立にとっては照合が困難であっても、個人情報保有するJR東日本にとっては照合が容易であるため、「照合は容易である」という判断が下されることになる。

ではどうすれば良いかという、個人情報保護法は、告知と承認という手続を要求している。しかし、改札で目的変更の度に告知と承認の手続を行うことは非現実的だ。実

は、現行個人情報保護法では、規約や目的を変更する際にどうすれば良いかという点についてのデザインが為されておらず、非常に大きな問題だ。この様な場合の対応を定めることが重要な世の中となっていくが、これは数年でできることではないだろう。

[問3]

既存のSuica利用者については、告知と承認の手続を踏まなければならないということになるのか？

[答3-1]

そういうことになる。

2014年11月に規約を改訂したカルチュア・コンビニエンス・クラブは、個人情報保護法のハードルは一応乗り越えた様に思える。それは、1ヶ月間、ウェブ上で規約改訂を告知し、(Yahoo! JAPANのIDを使用しなければならなかったという点で炎上はしたが)オプトアウトの方法を設けたからだ。この様な仕組みを最初の時点で作っておくということが重要である。

[答3-2]

個人の特性を不可能にすることはできない。個人情報保護法の改正が検討されているが、特定性を限りなく低減させたものであれば、かつ、一定の条件の下であれば、本人の許諾無しに利用することができる方向で検討が進められている。一定の条件とは、1)再特定しない旨の取り決め、2)提供先の明示、3)第三者機関への届出、等である。

情報提供後にデータがどのように利用されるのか、という点が非常に重要となる。情報提供によってダイレクトメールが届くだけということであれば、やはり反発を受けるだろう。一方、混雑していない電車を知らせてもらえたり、より安価にサービスを購入できるという本人のためになることであれば、受け入れられやすくなるだろう。サービスの設計次第であり、現在の世の中にあるもの、例えばポイント1%還元程度では納得されないのではないか。