

日本のデータ戦略は決して遅れてはいない——東大・越塚教授が語る期待と課題



東京大学 大学院情報学環・学際情報学府
教授

越塚 登

データ戦略を進めているのは企業や産業だけではない。“通貨”とも“21世紀の石油”とも喩えられるデータは、国にとってもますます重要度を増しつつある。日本政府が打ち出しているデータ戦略はどこまで進展しているのか、そこにはどのような課題があるのか。

「悲観することはない」と語るのは東京大学大学院 情報学環の越塚登教授。分野を超えたデータ流通プラットフォーム「DATA-EX」のコンセプトを提唱した越塚教授に、日本のデータ戦略の現状、課題、展望について聞いた。

取材／狩野英司（行政情報システム研究所）、平野隆朗（同）

文／末岡洋子

データ戦略では国は当事者であり、支援者ではない

データが重要な財であるということについては、もはや異論の余地はない。産業界ではデータを活用して競争力に繋げる動きがあるし、社会課題の解決でもデータが使われている。このような動きを受け、国もSociety 5.0を打ち出した第5期科学技術基本計画、そしてそれに続く第6期計画でもデータ戦略の重要性を指摘している。また、2021年6月には「包括的データ戦略」が閣議決定されている。

ここで、これまでの情報通信戦略の流れを振り返ってみたい。最初はマシンや光ケーブルなどのハードウェアが重要だった。その後、重要なレイヤーはソフトウェア、そしてサービスへと移り、現在一番上位にあるデータで競争力向上を図りはじめていると考えるといいだろう。かといって、ハードウェアやソフトウェアが重要ではない、ということではない。重要なものが増えていくと考えるべきだ。

このように、今や情報通信戦略というとき、メイ

ンのターゲットはデータになっている。公共の場も例外ではなく、データをどう使うかでサービスの質に差が出てくる。

特に日本固有の背景として、少子化が原因で進むと予想される縮小社会がある。その中で行政サービスはどうやって品質と規模を維持するか。現在の行政サービスにあるさまざまな無駄をなくし、最適化することが急務になっている。ここでデータは重要な役割を果たす。情報がデータになってはじめて、市民の手続きや業務を効率化できる。ゆくゆくは行政サービスの自動化へと繋がるだろう。

国はデータ戦略を進める上で、データ戦略はこれまでの政策とは次の2つの点で異なることを認識すべきだ。

1つ目は、データ戦略では国は当事者であるということ。日本で多くの基礎的なデータを保有している組織は国と自治体だ。これまでの情報通信政策を

データ戦略とベース・レジストリ

振り返ると、公共は補助金を出すなどの支援する側だったが、データ戦略ではデータを保有・提供するプロバイダであり、データを使うユーザでもある。産業界も、国・自治体がデータを公開することを待っている。国には、自らが主体になって、どんなデータを作るか、それを保守・運営していくかということが問われている。

2つ目として、意思決定のやり方をエビデンスベースに変えなければならないということ。EBPM

(証拠に基づく政策立案)だ。これまで行政においても、カンや過去の経験に基づいて意思決定をする場面も多かった。しかしデータ戦略の下では、エビデンスの実態であるデータを用いて合理的に判断をしたり政策を決定したりする、という考え方に変えなければならない。これは単にコンピュータを使うということではない。考え方をデータ中心に変えるということであり、このような意識を持っていない行政機関はまだ多い。

自治体の権限と標準化の調整をどう図るか

国はデータ戦略を進める上で、データを「使う」「揃える」「運用する」という3つを具体的に考える必要がある。

まずは「使う」から見てみたい。政府も自治体も、何らかの計画を立てるときは全て、エビデンスに基づいて決定しなければならない。これによって、きちんと効果のある施策を無駄のない形で打つことができるようになる。

次に「揃える」「運用する」だが、データに基づいてサービスを提供する上でデータが足りないのであれば、データを作成する必要がある。そこで、「包括的データ戦略」ではベース・レジストリの整備が入っており、コンセプト案が公開されている。

ここで問うべきこととして、新しい種類のデータを作るべきかどうか。これについては意見が分かれるところだが、新しいデータはなるべく作らず、既存のものが使えるならそれを使うという考え方が重要だと私は考える。データ活用にあたっての重要なポイントは、使い回せるかどうか。データを作っても、特定の分野で特定の用途に閉じた“サイロ化”の状態では、使い回しができない。ベース・レジストリを開発するに至った背景には、社会の基礎となるデータ、つまり幅広い分野で使えるデータが、使える形で整備されていないから作ろうという考えだ。従って、ベース・レジストリは使い回ししやすいものにしなければならない。

データ戦略を進める上で、もう一つ重要なのが体制だ。データは標準化なしには使えないが、標準化

をどうするか、運用をどうするかといったところを管理する組織や体制を行政の中に作らなければならない。現在、道路や橋の規格は国土交通省、薬の規格は厚生労働省、というように、それぞれの分野の担当省庁で規格化が進んでいるが、データについては体制をどうするのがまだ見えていない。米国ではアメリカ国立標準技術研究所(NIST)がこれらを行っているが、これに匹敵するような組織が日本にはない。

運用体制は、データの権限の関係もあり簡単には進められない部分だ。例えば、教育を見てみよう。大学の願書を電子化できれば便利だ。しかし、内申書が紙であるという事情などがあり、現時点では不可能だ。通知表のフォーマットを決める権限を持っているのは各学校であり、自治体でも文部科学省でもない。そのため、何らか工夫なしに、今のままで通知表を電子化すると、学校毎に個別のデータフォーマットとなる。

これまで日本の行政は、柔軟に動けるようにという狙いで、住民に近い自治体に権限を移行させてきた。しかしITから見ると、(日本の地方公共団体数である)1,788の国があるようなもので、統一システムを作ることは非常に難しい。ITで重要な標準化と、地域の柔軟性をどのように両立させるか。これも課題だ。

このように、ITそのもの、データそのものではなく、データがうまく使えて、効果が出せるように自治体の構造、国との関係を見直す必要がある。ITが

あることを前提としていなかったものを、どこまで見直ししてITがあることを前提にした構造にするか。正解はないし、すぐにできることではない。しかし、人口は確実に減っていく。行政も、現在のようには1,788種類も同じような行政システムを抱えることはできなくなっていく。せめて、将来的に見直しが必要で、どこが課題になるのかの洗い出しは進めていいのではないかと。そういう段階にきていると思う。



インタビューに答える越塚氏（AIS撮影）

動き出すベース・レジストリ構想、課題は？

そのような状況で進んでいるのが、ベース・レジストリ構想だ。ベース・レジストリは欧州で生まれたアイデアで、日本ではデジタル庁が中心で進めている。（図表1）

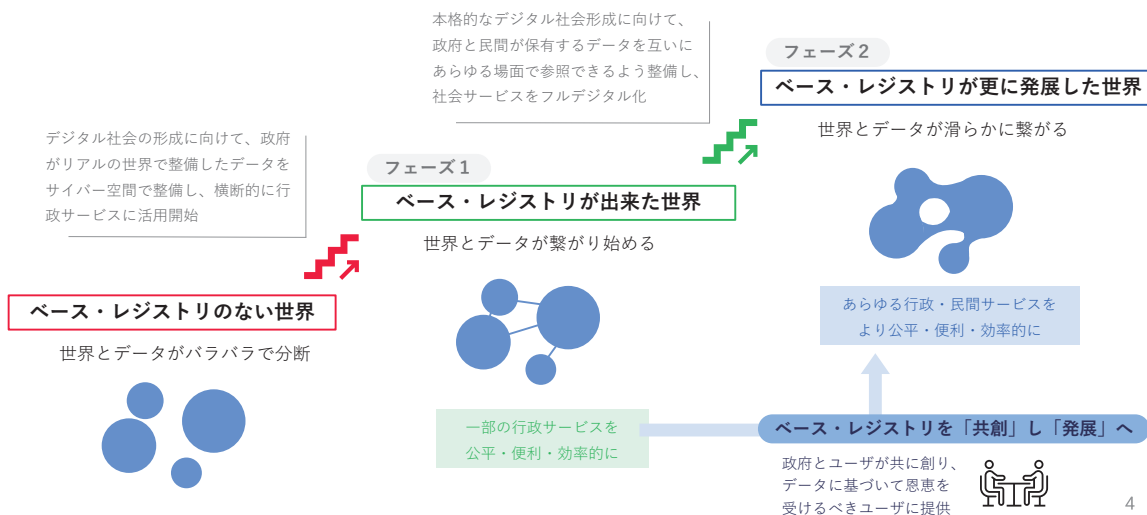
土台にあるのは、あらゆる人に利益になる日本の国のデータを整備するという考えだ。ある自治体にはデータがあるのに別の自治体にはないといったことが発生しないようにしようというもので、地図情報や郵便番号などの土地系、支援制度、法律・政令・省令などの行政系、法人の決算情報や事業所などの事業者系などの分野でデータを整備する計画になっ

ている。この中には、公開しても良いものもあれば公開できないものもあり、公開できないものについてはセキュリティを担保しておく必要がある。

例えば学校の事務手続きを考えると、自治体は、どこに・誰が住んでいるのかを把握しているので、来年4月に何人の児童が小学校に入学するかがシステムからわかる。だが自治体間でデータの形がバラバラで相互運用性がないと、引越して転校するときは一旦紙にする作業が生じざるを得ない。ベース・レジストリにより、このような問題が解決できると期待したい。

図表1 ベース・レジストリの世界観

ベース・レジストリで世界が変わる



（出典）データ戦略推進WG第1回資料より抜粋

データ戦略とベース・レジストリ

一方で、考えるべきことや課題も多い。最初に挙げたいのは、クオリティだ。具体的には、「データフォーマットのクオリティ」「データの正確性のクオリティ」の2つ。しかし、クオリティが高ければいいというものではない。クオリティの高さを求めすぎると、低いデータは公表しないということになる。ここでは、クオリティをきちんと評価し、高いクオリティのものは高いものとして、低いクオリティのものは低いことを前提に扱うように進めるべきだ。

関連して、トラスト（＝信頼性）も重要だ。トラストとは、このデータは信頼できるデータなのか、データにアクセスする人は信頼できる人なのかを知るためのものだ。誰かにデータを渡すときに、その人と主張している人が本人だとどうやって特定できるのか。現時点で一般的なデータ交換を行う際のトラスト確保のためにマイナンバーの枠組みは使えず、政府による本人確認ができる日本のIDはない。データ交換において、トラストは重要な要素だ。また、データの活用でよく使われる機械学習はデータの不備に弱いメカニズムで、偏ったデータでトレーニングすると偏った結果がでる。そういう意味においてもトラストは重要で、そのための仕組みを日本国内、そして諸外国と連携しながら作らなければならない。

トラストのための一つの策が、データの履歴を残すこと。そのデータを作ったのは誰か、どういう流れでここまできているのかという記録で、その履歴そのものも改ざんされていないことを電子署名やブロックチェーンなどで担保する。

2つ目の課題がリテラシーだ。これも2種類あり、デジタルそのもののリテラシー、そしてエビデンスベースで科学的に業務を進めたり意思決定したりするリテラシー、両方が必要だ。

データ中心の考え方が必要といわれるが、日本では情報系システムというとハードウェアから考えがちである。道具は揃っても、肝心のデータでどう表すのかといった部分が抜けていることがよくある。こういう場合、住民目線から考えると良いのではない。通知表の例だと、システム構成から考えるのではなく学生や保護者が通知表データをどう利用するかと考えるといいだろう。

この2つの課題からいえることは、制度作りにおいて、誤りがあることを前提としたアプローチをとるべきということだ。誤りがないことを前提に制度を作ると、うまくいかない。そうではなく、誤りがあつたときの例外処理をルール化しておく。間違えない縛りではなく、間違えたら修正するという縛りを与える。

各国が進むデータ流通プラットフォームの取り組み

クオリティとトラスト、そしてリテラシー以外の課題として、人材、国際連携についても触れておく。

人材については、外部人材に期待したいところだが、内部でも育てなければ追いつかない。デジタルに限ったことではないが、社会人も常に勉強しなければならない。デジタルの次にも新しいことはどんどん出てきて、その度にそのスキルがある人材が必要になる。働き盛りで最前線にいる人が勉強する以外の方法はない。そこを新しい人材や外部に丸投げすると、空洞化がおこり、組織はどんどんだめになる。

国際連携とは、データが国をまたいで使われるためには各国が協調し、共通化できる場所は共通化

すること。現在各国がデータ戦略を進めている背景として、データに規範やルールがないという現状がある。すでに欧州のGDPR（EU一般データ保護規則）など自分たちでルールを作ろうという動きが出てきているが、これが不健全な形で進むとローカル化が進み、国と国の間でのやりとりができなくなる。そうになると、データの潜在性を生かせない。

日本と欧州でデータに関する条約があり、欧州と米国でも条約があり、米国と日本でもあるという形をとると、拡大させていくことが難しくなる。ひょっとするとA国の企業はB国のサプライチェーンからデータをもらえないとか、C国の会社はD国の製造ラインに入れられない、などといったことが生じるかも

しれない。規範について合意できるところは合意して、世界共通部分を大きくしていくことが望ましい。

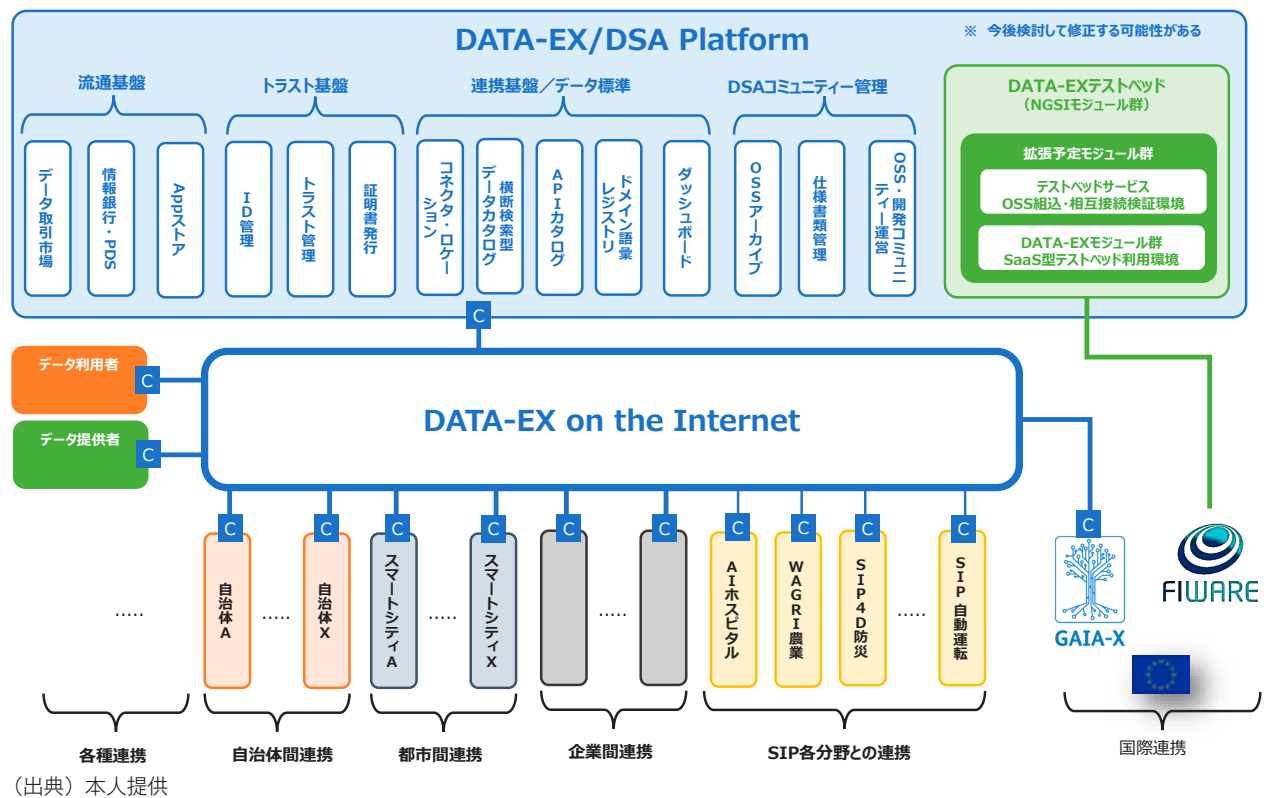
これらの課題から、私は「DATA-EX」のコンセプトを提唱した。各分野でデータの流通と利活用が進む中で、「DATA-EX」では分野を超えたデータ連携の実現を目指す。データ連携に係る既存の取り組みが協調する連邦型のプラットフォームで、分野の壁を越えてデータを容易に発見できるようにする。現在「DATA-EX」は、一般社団法人データ社会推進協議会（Data Society Alliance：DSA）が進めている。（図表2）

このようなデータ流通プラットフォームは、世界各国で動きがある。欧州にはGAIA-Xなどがあり、インドもIndia Stackという取り組みを進めて

いる。中国はデータプラットフォームを、データ取引市場として展開している。国によりポリシーやスタイルが異なるが、これらのデータ流通プラットフォームがどのように発展していくのかが、情報通信業界では熱い論争になっている。その意味では、“GAFAM”の次の時代に向け、協調的なプラットフォームをどのように作るのかという動きはすでに始まっている。

「DATA-EX」は具体的な実施段階に入りつつあるし、トラストも、ベース・レジストリも少しずつ進んでいる。日本は何もやっていないというわけではない。先頭を走っているわけではないかもしれないが、きちんと対策は打っていると考えている。

図表2 DATA-EXデータ連携サービスの将来展望と開発環境



時間がかかるデータ戦略、やり抜く覚悟を

これから国がデータ戦略を進めていくにあたって覚悟しておくべきことは、このような取り組みは時間がかかるということだ。加えて、ベース・レジストリやプラットフォームなどのバックエンドの整備は国民生活に直結しないので、その間、国民からは「デジタル化やデータ戦略で生活が良くなったの？」と、疑念の矢面に立たされる。そうなると、行政は国民の生活に直結する政策を優先したくなるかもしれない。また、これらの政策は1年で終わるようなものではないので、毎年こうした矢面に立つ覚悟で進めていかなければならない。

コロナ禍で行政のデジタル化の遅れが露呈したと

して、“デジタル敗戦”という言葉が聞かれる。だが私は、日本は悲観的な状況ではないと考えている。世界を見渡しても、デジタル化が全てできている国はなく、強い部分、弱い部分、それぞれある。日本は20年ぐらい前にExcelで一度DXをやった。それから時代が変わったに過ぎず、これから取り組みに着手することで日本でも一気に新しいデジタル化を進めることは可能だろう。

ネガティブに考える必要はない。着実にやっていけばできる力はあるので、長期的な視点を持って目の前の課題に取り組むことが大事だ。

越塚 登（こしづか のぼる）

専門は計算機科学（Computer Science）。特に、Ubiquitous ComputingやIoT（Internet of Things）やOpen Data、Operating System、Computer Network、Human Computer Interface、Block Chainなどの研究に取り組んできた。近年は情報通信技術を活用した社会課題の解決や、社会基盤としての情報システムに関心を持ち、特にスマートシティ（特に都市OS）やデータ流通プラットフォーム等の研究を中心に活動している。

現在、東京大学大学院情報学環教授、一般社団法人データ社会推進協議会（DSA）会長、一般社団法人オープン&ビッグデータ活用・地方創生推進機構 理事、高知県IoT推進アドバイザー、小田原市デジタル政策最高顧問等の活動を手がけ、デジタル庁 デジタル社会構想会議 委員、同 データ戦略推進WG 委員、内閣府スーパーシティ/スマートシティの相互運用性確保に関する検討会座長、東京都「ICT先進都市・東京のあり方懇談会」委員等の政府自治体委員を務めている。