

# ビッグデータ

## 企業におけるビッグデータ活用を考える

### 巻頭言

ビッグデータを新たなビジネス創出やマーケティングの拡大に活かしたいと考えている企業が増えている。一方、何から着手してよいか分からないという悩みを抱えている企業も多い。こうした状況に対して、先進的な事例や、ビッグデータにこれから取り組もうとしている企業が押さえておくべきポイントについて、『ビッグデータの衝撃』（2012年 東洋経済新報社）の著者でもあり、先端ICTの動向に詳しい、株式会社野村総合研究所 上級研究員 城田真琴氏にお話を伺った。



城田 真琴 氏

株式会社野村総合研究所 情報技術本部  
先端ITイノベーション部 上級研究員

大手メーカーのシステムコンサルティング部門を経て、2001年より現職。ITアナリストとして先端テクノロジーの動向調査、ベンダー戦略の分析、国内外企業のICT活用調査を推進。同時に、ICTの将来予測とベンダー、ユーザー双方に対する提言を行っている。ITホワイトボックス（NHK）などのテレビ出演、講演など多数。

### 言葉の浸透は進んだが、 実際の取り組みは

日本では、「ビッグデータ」は2012年から注目され始め、言葉自体の認知度は非常に高まりました。野村総合研究所がユーザー企業を対象として継続的に実施しているアンケート調査の結果からも、2013年9月の時点で、「ビッグデータという言葉を知らない」と回答した情報システム部門の方は、全体の3%しかいませんでした。我々のところへも「ビッグデータで何かできないか」という話がくるようになりましたし、ビッグデータの活用戦略を立てる企業は少しずつ増えてきています。

しかし、実際に取り組みを始めているかといえは、その割合は諸外国に比べると圧倒的に少ないですね。先ほどのアンケート調査は、日本を含む世界5か国で実施していますが、日本は「すでに取り組んでいる」4.8%、「試験的に取り組んでいる」4.4%と、合わせても10%に満たない状況です。一方、アメリカは19.6%と16.3%、イギリスは15.6%と16.3%、ドイツは15.0%と11.7%、そして中国は最も進んでいて、20.6%と36.3%を占めています。

このように、日本だけがビッグデータの取り組みが遅れている原因としては、1つに、「データサイエンティスト<sup>\*1</sup>」と言われ

る人材の不足、もう1つは、鉄道会社による乗降履歴の第三者提供の一件に代表される問題です。個人情報に関わる問題が社会的にクローズアップされてしまったことで、企業が顧客情報の活用をリスクと感じて萎縮していると考えられます。

### データサイエンティスト不足は、 チームで対応を

データサイエンティストには広範なスキルが求められます。その1つは、エンジニアとしてのスキル。Hadoop（ハドゥープ）<sup>\*2</sup>と言われる大規模データの分散処理フレームワークなどを駆使してデータベースを操作したり、コーディングを行うスキルです。次に、統計解析、数学的な知識。大学で統計解析ソフトウェア「SPSS」や「SAS」などを使ったことがある人です。そして、コミュニケーション能力。一昔前にデータアナリストと呼ばれた人達のイメージは、密室にこもってデータ分析に没頭している、というものですが、データサイエンティストは、意思決定者に分析結果を分かりやすく伝え、「次にこういうことをすべきですよ」とアドバイスを行い、事業部門が実際の行動に移せるところまで寄り添う。そこまでのスキルが求められます。単純

\*1 データサイエンティスト:

データ活用の専門家。統計解析や機械学習、分散処理技術などを用いてデータを分析し、企業の課題を的確に把握し、解決策を提案する。本誌10ページ「技術解説」参照

\*2 Hadoop（ハドゥープ）:

本誌10ページ「技術解説」参照

にレポートを書いて終わり、という人は、データサイエンティストとは呼ばれません。

分析結果を分かりやすく伝えるという点では、データを見る化する、いわゆる「ビジュアライゼーション」のスキルも重要です。単純な数値の羅列だけを見せられた経営者が、レポートを読まずに終わってしまうケースは多々あります。多忙な経営者が1分で分かるくらいに、情報を見やすく整理するビジュアライゼーションのスキルは、日本ではまだ注目されていませんが、アメリカなどでは重要視されています。

そしてもう1つ、業務知識も求められます。ビッグデータ活用の最終ゴールがビジネスにつなげることなら、そのビジネスを分かっていると、どういう仮説を立てればいいのか、どういうデータを持ってくればいいのか、の勘所をつかめません。業務知識は、言わば土台となる部分ですから、そこが抜け落ちたデータサイエンティストには、十分な成果を期待できません。業務知識を持ち、なおかつエンジニアとしてのスキルや統計解析のスキルを備えてこそ、ビジネスに役立つ新たな知見などを導き出すことが期待できます。

とはいえ、これらのスキルを全て兼ね備えた、スーパーマンのような人材は、そう簡単には見つかりません。そのためにビッグデータへの取り組み自体を断念してしまうことが日本企業には多いのですが、実は、アメリカでもデータサイエンティストと呼べる人は一握りです。そして、その一握りの人材は、GoogleやAmazon、Facebookといったネット系の世界的大企業が高給で採用し、一般企業にはまわってきません。そこで最近では、一人のスーパーマンではなく、それぞれのスキルを持った人を集めてチームで対応していきましょう、ということが言われています。その場合に必要となるのが、通常のシステム開発でいうところのプロジェクトマネージャーです。企画・提案、コンサルティングができ、かつ人材管理もできる人のもと、チームとしてデータサイエンティストの業務を遂行するということです。

社内の人材だけでは実現できない場合は、コンサルティング会社やSIerの人材を活用する方法もあります。ただ、外部の人間が本当に自社の業務を100%理解してくれていればい

いのですが、必ずしもそうとは限らないので、常駐してもらって、実際の業務を見ながら分析モデルを作ってもらい、というパターンが多いようです。また、たとえば、EMC ジャパンが提供している5日間のトレーニングコースのような、ベンダー主催のデータサイエンティスト研修を利用して、社員に基礎知識を習得させようとしている企業もあります。外部研修で基礎を作ったうえで、常駐の外部コンサルタントと並走しながら実践を積んでいく、といった方法でデータサイエンティストを育成している企業もあります。ないものねだりをしては仕方ありませんから、こういう取り組みは現実的と言えます。

### 個人情報情報を正確に理解し、「透明性の確保」と「本人関与の機会」を

乗降履歴の第三者提供の一件は、何が問題だったかという、第三者提供に対して利用者に何の通知も告知も行わず、勝手に販売しようとしたことです。個人情報を取り扱う際の基本的な考え方は、どういう個人情報を取得して、それをどういう目的で誰に渡すかを透明化する、ということです。また、問題化した後、鉄道会社は、収集したデータの氏名や住所などを削って匿名化したうえで第三者に提供していると説明しましたが、乗降客の少ない駅の利用者のデータが一定期間蓄積されると個人が特定できてしまう可能性があり、個人情報となり得るものです。単純に氏名・住所を削除さえすれば個人情報でなくなるというのは認識が甘いといわれても仕方ないでしょう。問題はもう1つあって、「乗降履歴を第三者に提供されたくない場合はここに電話してください」といったオプトアウトの手段が、問題化するまで準備されていませんでした。

「透明性の確保」と「本人関与の機会」は、個人情報を取り扱う際に必須です。総務省が主催し、2012年11月から2013年6月にかけて開催された「パーソナルデータの利用・流通に関する研究会」の報告書が公開されていますが、それに沿っていれば、この一件は基本的には問題ない話でした。きちんと考慮しないで事業者側で先走ってしまった、というのがこの一件の大きな問題でした。

※3 オプトアウト:

企業が個人の情報を収集・利用・第三者提供する場合、後から情報を削除する機会を提供すること。一方、事前に本人の許可を得ることは「オプトイン」という。

※4 パーソナルデータの利用・流通に関する研究会:

総務省「パーソナルデータの利用・流通に関する研究会」報告書  
http://www.soumu.go.jp/menu\_news/s-news/01ryutsu02\_02000071.html

## 特集 ビッグデータ

巻頭言	企業におけるビッグデータ活用を考える	2
事例	浜松市様／沼津工場 Akisai	6
技術解説	知っておきたい！注目のビッグデータ技術	10
富士通の取り組み	ビッグデータ活用を成功に導く 目的の明確化からお客様を支援	12

## C O N T E N T S

● 講演録	14
株式会社エスワイフード 代表取締役会長 山本 重雄 氏	
● 豊かに生きる誌上セミナー HUMAN HUMAN	16
株式会社 MANY ABILITIES 代表取締役 野原 秀樹 氏	
● Family's Information ほか	17

第三者提供が問題化した事例は、海外にもあります。オランダのカーナビメーカーは、利用者の車の速度情報や位置情報を、カーナビの渋滞情報の提供に活用するとともに、それらの情報を道路計画の策定に活用したい政府や自治体に販売していました。そこまではよかったのですが、あるとき、警察にも販売していることが表面化しました。警察は、その情報からスピード違反の多い場所を特定できますから、その場所に速度違反取り締まり装置（通称オービス）を設置すれば、検挙率が上がります。これが利用者の反発を買った結果、メーカーは謝罪に追い込まれ、警察への販売を取りやめることはもちろん、利用規約の改訂も行われました。

利用者に対し、何のメリットも用意せずに個人情報を収集することは、仮に法律上は問題ないとしても、利用者の心情的には受け入れにくく、ビジネスとしては難しくなります。その点、利用者に分かりやすいメリットを提供することで、うまく個人情報を収集している例として、アメリカ・フロリダ州のウォルト・ディズニー・ワールドが2013年1月から試行している「MyMagic+」というサービスがあります。RFID タグを埋め込んだ「マジックバンド」と呼ばれるリストバンドに、ルームキーやテーマパークチケット、食事やお土産を購入する際のクレジットカードの機能を持たせ、一元管理しようというものです。事前に Web から個人情報や好きなキャラクターを登録しておく、そのキャラクターがグリーティングをしてくれたり、誕生日を祝ってくれたりします。子どもが喜ぶますから、親は進んで子どもの名前や誕生日といった個人情報を提供するわけです。マジックバンドによってディズニーは、来園者がどんなアトラクションを利用したか、どのキャラクターと接触したか、何を買ったのかといった情報を収集し、より個々の来園者にカスタマイズしたサービスの提供が可能になります。もちろん、取得する情報の内容、収集の目的、オプトアウトの方法、第三者提供の有無などについては、利用者が Web から登録する際に明記しています。

## 日本における注目のビッグデータ企業



日本におけるビッグデータのキープレイヤーは、“ビッグデータならではの”ということでは、データ分析の専門企業でしょう。多くのデータサイエンティストを抱え、企業から預かったデータを分析してその結果を返す、ということを行っていて、その代表がブレインパッドです。ビッグデータブームにも乗って、大手企業との協業や合併会社設

立などを行い、ベンチャー企業が今や東証1部上場まで果たしました。他に、ALBERT（アルベルト）、iAnalysis（アイアナリシス）といったデータ分析の専門企業があります。

とはいえ、データそのものがないと始まりませんから、強いのは、オリジナルのデータを持っている企業で、さらに分析できる人材を社内に抱えていれば最強です。そういう企業の多くはネット企業で、楽天やリクルート、サイバーエージェント、クックパッドなどが代表的です。ネット以外の企業で目立っているのは、大阪ガスです。「ビジネスアナリシスセンター」を情報通信部内に設置し、10人ほどのデータサイエンティストを抱えています。

ただし、自社のオリジナルデータを活用している企業の多くは、ビッグデータなどと言われる前からデータ分析に取り組んで成果を出していて、それがビッグデータブームで注目されている、というだけのことです。Amazon のレコメンド機能にしても、登場したのはビッグデータがブームになる前ですし、こうした企業のビジネスが、現在も引き続き順調、ということでしょう。

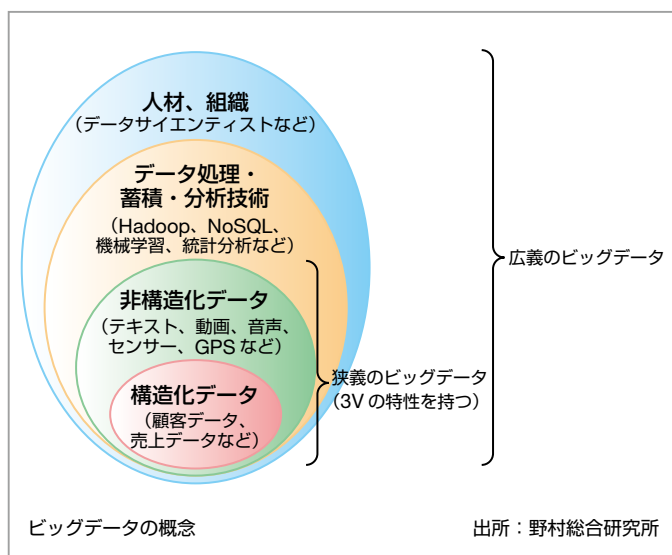
ビッグデータブームの背景には、コンピューティング・パワーの向上があります。以前は、大量のデータを蓄積したくてもままだハードディスクが高価でしたから、一定以上のデータは破棄せざるをえませんでしたし、大量のデータを持っていても分析に1か月かかっていたは、ビジネスのスピードに合いません。それが最近になって、処理能力が向上し、ハードディスクも安価になりました。しかも、クラウドサービスを利用すれば、自前で何千万、何億円ものデータウェアハウスを構築しなくてもよくなり、ユーザー企業がデータ分析に取り組むための敷居が低くなってきています。

## どのようなデータも“お宝”に変わるのか

ビッグデータの定義は、一般に、「既存の一般的な技術では管理するのが困難な大量のデータ群」であり、その特性は「Volume（ボリューム：量）」「Variety（バラエティ：多様性）」「Velocity（ベロシティ：更新頻度・発生頻度）」という、Vで始まる3つのキーワードで示されます。しかし、これは狭義のビッグデータであって、現在のビッグデータを取り巻く状況を含めた広義のビッグデータは、「3Vの面で管理が困難なデータ、および、それらを蓄積・処理・分析するための技術、さらに、それらのデータを分析し、有用な意味や洞察を引き出せる人材や組織を含む包括的な概念である」と考えられます。

ただ、ビッグデータは、定義はそれとして、本質的にはそれによって、何かにあるいは誰かに、きちんとメリットを出せるかどうか、というものだと思います。

ビッグデータに対してありがちなのは、「なんとなくたまって



いるこのデータで何かできないか」と、漠然とした甘い期待を抱くケースです。「優秀なデータサイエンティストなら、こんなデータでもお宝に変えてくれるだろう」と思っている人がいますが、それはさすがにないでしょう。そもそも、目的が決まらないうちに、収集すべきデータも決まりませんから、ビジネスで結果を出すには道のりが遠くなってしまいます。

また、収集したデータを分析する前に、データの中にあるゴミを取り除く、いわゆる“クレンジング”という処理が必要になります。こうした作業は、顧客データの場合はマーケティング部門がこの処理を担当することもあるようですが、多いのは情報システム部門が担当するケースです。全行程の7割くらいが、こうした前処理に費やされるという話もよく聞きます。データ分析よりも、その前処理に多くの作業負荷がかかることで、くじけてしまう企業もあります。「意外に時間も金もかかる。ぱっとやっばつといい結果が出ると思っていたのに」と、上層部の人ほど感じるようです。

ビッグデータ活用を、社内のどの部門が推進していくべきか。これは、ケースバイケースだと思いますが、日本で聞いてみると、「情報システム部門の仕事」と回答する企業が多いですね。しかしながら、本質的には、ビッグデータのオーナーは事業部門ではないでしょうか。ビッグデータ活用によって誰が最終的な受益者になるかという点、業務を分かっている事業部門になるわけですから、彼らがオーナーになって、「我々はこういうゴールを目指している。それをデータドリブン（データ駆動型）で実現できないか」という相談をデータサイエンティストに打ちかける、というのが、適切なやり方だと思います。

## ビッグデータだからといって 身構える必要はない

ビッグデータを活用するには、まずは目的をはっきりさせること、と先ほど申しましたが、「そのハードルが高い」と考え

ているなら、ビッグデータだからといって身構える必要はありません。「自社の課題は何か」「課題の解決にデータが使えないか」と考えて取り組めばいいと思います。いくつかそういった事例を紹介します。

国内では、コンビニエンスストアを中心とした共通ポイントカードによるID-POSデータ（顧客の属性情報を含むPOSデータ）の分析が多く行われるようになってきました。以前は、購入者の属性を把握するために、たとえば、「女性20代」といった情報を、レジ担当者が目で見てレジに打ち込んでいました。ところが、忙しい時間帯になると、目的意識の低いアルバイト店員などは、自分が押しやすいキーを押しがちなことから、正確な属性が得られないという課題がありました。それが今では、IDカードで決済してもらうことで、正確な属性を把握できるうえに、購入者の居住エリアや、その人がリピーターなのか、たまたま通っただけなのか、といったことも把握できるようになりました。

ソフトバンクは、パケ詰まり（通信速度が極端に遅くなる現象）の解消にビッグデータを活用しています。利用者のスマートフォンにインストールされたアプリの通信ログから、どの場所での時間帯にパケ詰まりが多いかを詳細に把握して基地局の設置計画に役立て、「つながりやすさNo.1」を実現しました。

ホンダの場合は、同社の「インターナビ」により、数秒間隔で取得した車の位置情報をもとに急ブレーキの発生箇所を特定した情報を埼玉県に提供しています。実際に急ブレーキの多い場所へ行ってみると見通しが悪いことが分かり、その原因となっている大きな街路樹を剪定することで事故の発生件数が減った、といったように、地域の安全活動に貢献しています。

日本におけるビッグデータ活用の先駆けともいえるコマツは、GPSやセンサーを使って建設機械の稼働状況を遠隔監視する「KOMTRAX（コムトラックス）」によって、保守サービスの効率化や、顧客に対する適切な使用方法のアドバイスなどを行っています。初期バージョンが登場したのは1999年ですが、現在のビッグデータ活用のエッセンスが集約されており、その手法は、製造業を中心に非常に参考になると思います。

コマツのような、機械にセンサーを取り付けて故障を予兆検知するというビッグデータ活用は、今後普及していくと考えられます。実際、取り組みたいと考える企業も多いですね。日本はネットの接続料金が安いので、常に機械と通信して挙動を監視するといったことが、比較的始めやすい環境にあります。また、故障する前に部品交換できれば、アフターサービスの一貫にもなりますし、リプレースのご提案もしやすくなります。顧客との接点を常に持つておくという意味でも有用です。このように、今の業務の延長上で新たに提供できるサービスは何か、ということからビッグデータに取り組むのもいいでしょう。

■ 浜松市様

# SNS 上の市政への声と市民意識調査結果をデータ分析 可視化することで30年後の市政イメージを描き出す



次世代に責任をもてる総合政策として、30年先を見据えた市像の策定を進めている浜松市様。市民インタビューのデータに加え、ソーシャルメディア上のビッグデータを合わせて分析し、市民が潜在的にもっている期待や問題意識、市外から見た同市の印象を引き出している。行政発想の延長線に施策を求めるのではなく、行政の既存常識の壁を打ち破る政策を求める道筋を見出しつつある。

2013年、浜松市様は30年先の未来の姿を見据えた新・総合計画「(仮)浜松市未来ビジョン」を策定すると発表。自治体の総合計画スパンとしては非常に長期的な30年という期間もさることながら、策定に向けた基礎調査にビッグデータ活用を取り入れたことで話題を呼んでいる。2015年度を起点に30年先を見据え、「世代を通じて共感できる『未来』を創造」する。企画調整部次長 企画調整部企画課課長の川嶋朗夫氏は、「5年、10年後の計画策定となると、財政事情から制約が出てくるなど生々しい話になりますが、30年後となれば『そもそも、どうあるべきか』の観点で考えることができます。親と子の年齢差が30歳程度であり、30年は次代をイメージしやすいという発想もありました」と説明する。

「(仮)浜松市未来ビジョン」は、市長、有識者、地域企業関係者、学生を含む公募委員からなる策定委員会を中心に議論され、2014年12月末に策定される予定だ。その基礎調査のひとつとして「ビッグデータを活用した市民意識調査」がある。インターネット上に流れる浜松市に関する声や市民アンケート、市民インタビューのテキストデータ、市民の声システムデータを活用し、今後の市政に求められる市民の期待を抽出する。

なぜビッグデータの活用なのか。背景のひとつに市民意識調査の難しさがある。市は従来、アンケートなどにより定期的

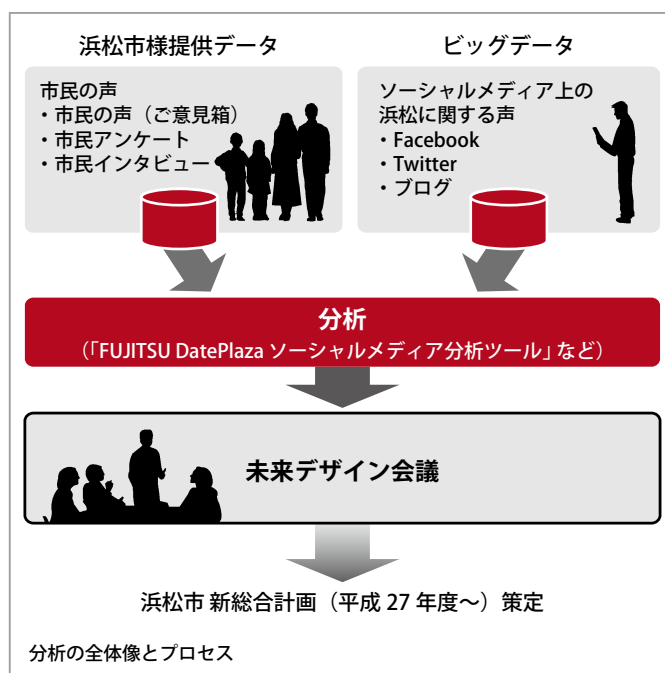
に意識調査を実施、結果を蓄積してきた。しかし「福祉行政に満足していますか？5段階で評価してください」といった定型的で一方的な質問への回答は、必ずしも市民が語りたいことではない。30年後の理想の姿を描くプロジェクトでは、より多面的に意識を浮き彫りにする必要があるが、多数の市民に、市が担う福祉、土木、教育など全ての事業に関する評価や意識を聞き出す調査は難しい。

これらの課題解決の突破口と目されるのが、膨大なテキストデータを分析する「テキストマイニング」だ。第1段階として、浜松市に対する要望を聞いた市民136人へのインタビュー結果をテキスト化し、A3用紙1枚に収まる「概要マップ」にまとめた。どのような単語がよく登場しているか、各単語の関係性はどうかを分析することで、市民が浜松市に期待していることや課題を可視化できたという。「テキストマイニング」は決して新しい手法ではありませんが、概要マップを作成したことで、行政に対する市民の認識と私も行政側の認識に異なっている点があることが示されました。例えば、行政は農業を産業と位置づけていますが、市民は農業をもっと特別なことと考えています。行政からみて想定外の発想は、実績に基づいて数年後を予測するフォアキャスティング式政策の限界を打ち破るカギになります。さらに30年後の浜松市のあるべき姿を描き、そのうえで

今そして10年後どうあるべきかの発想を導くバックキャストリング式政策を具現化してくれると期待しています」(川嶋氏)。

今後、「(仮)浜松市未来ビジョン」策定のためのビッグデータ活用は新たなステージに入る。企画調整部企画課の金子哲也氏は、次の一歩についてこう語る。「Web上で発信された浜松市の未来に関連するビッグデータを分析することで、思いもよらない斬新な発想が出てくるのは間違いないでしょう。しかし斬新なアイデアを活かすには、これまでのデータと結びつけて、どう分析するかが重要だと考えています。既存のデータとビッグデータをうまく合わせていくことで30年後の未来像に一歩一歩と近づいていけるだろうと実感しています」。

発想の壁を打ち破るビッグデータ分析は、既存常識や成功体験にとらわれない商品・サービスの開発などの有効手法ともいえそうだ。



浜松市様 <http://www.city.hamamatsu.shizuoka.jp/>  
静岡県西部に位置し、県内最大面積。12市町村の広域合併により、2007年、中核市から静岡県内で最大人口80万人を擁する政令指定都市に移行した。

## 地域振興や産業振興に向けた活用事例

地域振興において、行政の果たす役割は大きい。浜松市も収集したデータをオープンにすることで、地域の企業が新しい事業の芽を見つけるきっかけにしたいと考えている。行政が従来になかった視点でデータを活用することは、企業を活性化させ、市民生活をより豊かにすることを可能にする。限られた資源、労働力を最大限に利活用し、地方の力を持続的に成長していくためにも、官・民が一体となってビッグデータを活用していくことが求められている。

### ■位置情報を活用した被災地観光支援プロジェクト

KDDI株式会社と株式会社コロプラは、観光客の「位置情報ビッグデータ」を活用した観光動態調査レポートを、東日本大震災の被災県、岩手、宮城、福島に提供している。「位置情報ビッグデータ」は、KDDIがスマートフォン利用者から同意を得て個人情報取得したうえで、個人を特定できないように加工した性別・年齢層データと、位置情報データの履歴。これらをコロプラ社に委託し、特定の観光地域における来訪者、旅程、立ち寄り先、宿泊地、交通手段などの調査レポートを作成し3県に提供している。位置情報は、各観光エリアを一定の大きさの区画(地域メッシュ)に加工されたものを利用。例えば、基地局の緯度・経度と基地局への接続時間を元データに、観光エリアごとの時間帯別流入者数を算出し、来訪者の性別・年齢層などのデータとあわせて分析することで、滞在時間を延ばすための具体的な施策検討を可能にする。年齢層や交通手段などのターゲットを絞ることで、効果的な誘客施策の実施など、観光振興計画の策定や地域経済活性化に向けた施策に活用されることが期待されている。

### ■企業取引ビッグデータによる被災地産業復興加速策

被災地の産業復興は道半ばとされるが、企業取引ビッグデータを活用して被災地域の復興を加速、さらには国や自治体にとっての戦略的産業振興策の立案に活かす道が開けてきた。その可能性を示唆するのが、株式会社帝国データバンクが保有する被災地所在の12万社を含む約70万社の企業データと450万件におよぶ取引データ、そしてこれらの分析結果である。試みでは、被災地企業の取引先の取引先、さらにその先の取引先へと広がる取引ネットワークを分析。その結果、取引のつながり目となる「コネクターハブ企業」の存在が明らかとなり、同企業を復興の担い手として活用することが、被災地産業の復興を加速させる方策につながるのではなどの知見が得られているという。また、調査分析対象を一般企業に向け、全国規模に広げれば、企業の経営戦略の立案と実行に有効ではないか。あるいは国や自治体の蓄積する膨大なデータとの相関関係を分析することで、より有効な戦略的な産業振興策が立案できるのではないかと、期待を集め始めている。

■ **沼津工場 Akisai**

**24時間365日、栽培環境情報を収集  
蓄積されたデータから気づきを得る**

自然を相手にする農作物づくりは ICT の活用が難しい分野の 1 つだ。現在、富士通沼津工場内の Akisai 農場では、生産現場の見える化の研究が進められている。センサーや Web カメラを使って 24 時間 365 日、栽培環境を監視し情報を収集。蓄積された多様かつ大量のデータから新たな気づきを生み出していく。また見える化により得られたノウハウの栽培テンプレート化も検討中だ。クラウドとビッグデータの活用による農業の新しい時代が幕を開けようとしている。

■ **農業の見える化に取り組む Akisai 農場**

おいしい野菜や果物には農業生産者が長年培ったノウハウが活きている。勘や経験が大切となる農業の生産現場でどのように ICT を活用できるのか。2008 年頃から富士通は国内の農業生産者と ICT 利用の実証実験を行ってきた。

「農業生産者の勘や経験を数値化し、農業の生産現場を見える化する事は日本の農業に新たな可能性を拓きます。ICT を活用する農業の実現に欠かせないのが、いつでもどこでもサービスを利用できるクラウドと、多様かつ膨大な情報を扱うことができるビッグデータの融合です。富士通沼津工場内にある Akisai 農場では、実際に作物を育てながら農業の見える化の研究に取り組んでいます」とイノベーションビジネス推進本部 秋野氏は語る。

Akisai 農場ではハウス 2 棟による施設栽培と露地栽培で ICT が活用されている。例えば施設栽培では、24 時間 365 日、センサーと Web カメラにより温度、湿度、二酸化炭素、日照量、生育状況など栽培環境をクラウド経由で監視、計測。計測結果はデータとして蓄積され、グラフなどに視覚化することで気づきを促す。また施設では取得した栽培環境データに基づき、



天窓、換気扇など施設内の環境を ICT で自動制御する。

Akisai 農場における野菜の栽培は 2013 年 6 月からスタート。まだ数カ月しか経過していないが、すでに成果が見え始めている。その 1 つが、仙寿菜プロジェクトである。

■ **なぜ葉の色が赤く染まらないのか**

仙寿菜は熱帯アジア原産のヒユ科の植物、アマランサス。特に赤色の品種は高い抗酸化活性を有し栄養価も高い。仙寿菜プロジェクトは、夏の期間だけ露地で栽培されている仙寿菜の通年、水耕栽培に取り組み、栽培データを活用し最適な栽培環境の実現を目指す。同プロジェクトは、Akisai 農場の施設を使って国立大学法人岐阜大学、WSB グループと富士通が共同で行っている。

Akisai 農場における仙寿菜栽培で課題となったのは、葉が緑色から赤色になかなか染まらなかったことだ。「どういう条件なら葉が赤く染まるのか。8 月から 9 月にかけて 24 時間の温度や湿度の変化をグラフ化し、赤く染まってきた日と緑色に戻ってしまった日を見比べていたところ、温度の変化が葉の色に大きく影響することがわかりました。過去のデータと作業の結果を比較し、新たな気づきを得られるところが大きなメリットですね」とイノベーションビジネス推進本部 稲永氏は話す。

作物栽培の重要なパラメーターとして、温度は病害虫の発生や霜対策にも有効だ。また、空気中にとどれだけ水蒸気の入る余地があるかを示す飽差 (ほうさ) は食物の光合成との関わりが深い。毎日の平均温度を足していく積算温度はトマトやナスなど果実を食する野菜の収穫時期に大きく影響する。

飽差や積算温度などは、従来、農業生産者が作物栽培を通じて感覚的に知っていたことだ。作物栽培における匠のノウ

ハウに見える化する取り組みも始まっている。「富士通ではミカン農家と共同で栽培環境情報と作業情報をチェックしながら糖度の高いミカンの生産効率を高める試みを行っています。またブドウ園農家とは、栽培データの収集、分析を行い、ブドウの最適収穫時期の判断や害虫の発生予測に活用する取り組みを行っています」(稲永氏)。

今後の展望について「大学や農家などとの共同研究を重ね、様々な栽培データから導き出されたノウハウを

CO<sub>2</sub>センサー



Akisai農場の一角にあるハウスでは、仙寿菜が栽培されている



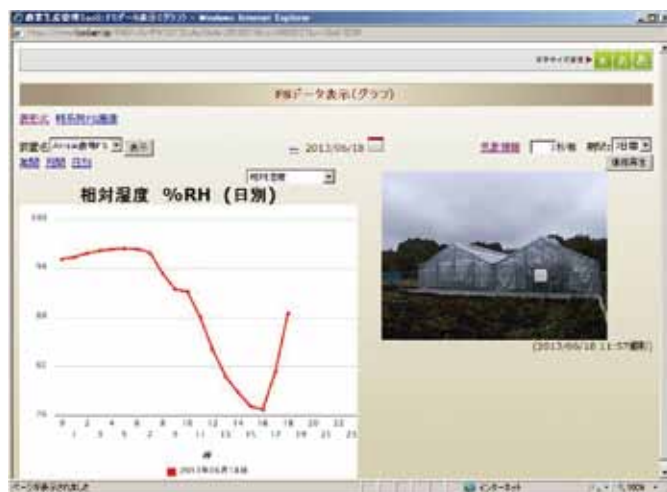
圃場センサー



Webカメラ

栽培テンプレート化しサービス展開していきたいと考えています。また、気象データや市況データなど様々なデータの活用も重要なテーマです。食・農クラウド「Akisai」を通じ、生産現場でのICT活用を起点に流通・地域・消費者をバリューチェーンで結び、日本の農業の発展に貢献していきます」と秋野氏は語る。

日本の農業も国際競争力が問われる時代。農業における生産現場の見える化は、日本の農業に新たな可能性を拓き、豊かな食の未来の実現に欠かせないものとなるだろう。



24時間365日、栽培環境情報を計測し蓄積データを活用

## センサーを利用した活動事例

### ■世界中の建設機械への個別サービスと経営判断情報をもたらすビッグデータ活用

建設機械メーカー大手のコマツ(株式会社小松製作所)は、グローバル展開する車両管理システムにおいてビッグデータを活用している。

「KOMTRAX(コムトラックス=Komatsu Machine Tracking System)」と呼ばれるこの車両管理システムは、顧客に納入した建設機械などに装着したセンサーやGPS装置からのデータを、携帯電話基地局や通信衛星、インターネットを介して同社サーバに集積し、分析している。データは、車両の位置、運転時間、冷却水、エンジンオイルの状況、燃料消費量など。顧客は、メンテナンス期日、車両ごとの月間稼働状況、省エネ運転状況、そして故障の予兆などを知ることができ、運用コストの低減の提案やスピーディーな修理を受けられるといったメリットを享受できる。また万一車両が盗難にあった場合にも、遠隔でエンジンを停止させる、盗難防止機能も利用できる。

その一方、「KOMTRAX」は、コマツにも様々な有用情報をもたらす。最たるものは建設機械の国や地域ごとの稼働状況だろう。これら情報からいち早く市場動向を予測し、営業強化策を打ち、場合によっては在庫調整の判断を下すこともできるのだ。実

際にコマツは、2001年から「KOMTRAX」を標準装備サービスとして提供開始し、リーマンショックによる世界的景気低迷も乗り越え、営業利益を確保し続けている。サービスは無償で提供され、国内では6万2,000台(2011年)、海外では30万台(2013年)の装備車両が稼働しているようだ。

ビッグデータを活用した「KOMTRAX」が、顧客とコマツにとってWIN-WINのソリューションとなり得たのは、次の3つの要素からだといえる。第1に挙げられるのが、世界各国で稼働するコマツ製建設機械の1台1台から24時間365日、刻々と生成される多種多様なデータを計測しデータ化するセンサー技術。人手に頼ってはいはなしえなかった「個別マーケティング」といえる。第2がデータの集積、処理、分析を安全かつ低コストで行うステージがクラウド環境で構築可能になったこと。そして第3が、膨大なビッグデータから顧客が真に求める情報を抽出し、具体的な施策に結びつける、データ分析力とサービス構築力の融合だ。

社会には、GPSや温度・湿度センサーをはじめ、各種センサーが設置され情報を収集している。センシングデータを活用した取り組みは、様々な可能性を秘めている。



## 知っておきたい! 注目のビッグデータ技術

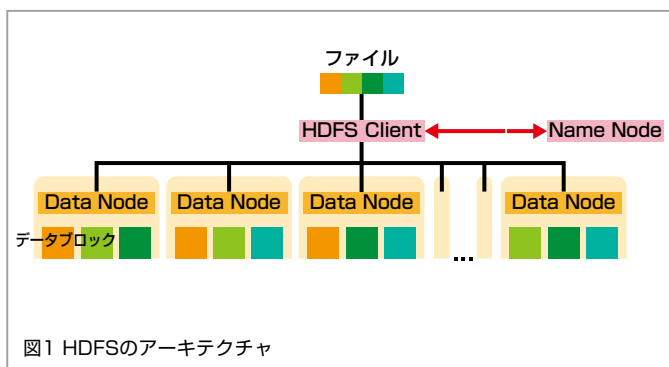
「ビッグデータ」という言葉は、2000年代半ば、スーパーコンピュータを必要とするような一部の研究開発分野で誕生した。以来、クラウドの進展やコンピュータの性能向上などにより、従来は困難とされていた大量のデータ群の蓄積・分析を可能にする環境が整ってきている。今回は、ビッグデータ基盤として注目されている技術を紹介するとともに、データサイエンティストについても触れる。

### 大量の非構造化データの高速処理を可能にした「Hadoop」

Hadoop (ハドゥープ) とは、Google が自社向けに開発した「Google File System (GFS)」と「MapReduce」をもとに、Yahoo! の技術者が中心となって開発した Java ベースの分散処理フレームワークである。大量の非構造化データを汎用 PC 上で高速に処理でき、またスケールアウト (サーバ台数を増やすことで処理能力を向上させる) を行いやすい点が注目を浴びている。現在は Apache ソフトウェア財団のプロジェクトとして、Apache ライセンスのもと、オープンソースとして公開されている。

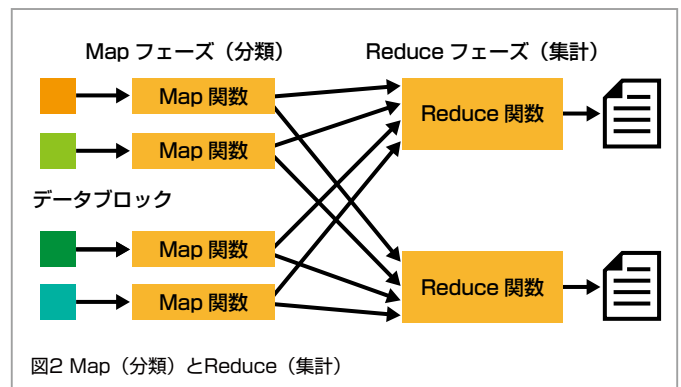
Hadoop の構成は、分散ファイルシステム「Hadoop Distributed File System (以下 HDFS)」と、その上で動作する分散プログラミングモデル「Hadoop MapReduce (以下 MapReduce)」から成る。

HDFS は、ファイルを均等な大きさのブロックに分割し、そのファイル名や保管ノード (サーバ) を一元管理する NameNode と、ブロックを実際に保持しデータを処理する複数の DataNode で構成される。1 つのブロックは複数の DataNode に保管され、これによって信頼性を向上させている (図 1)。



MapReduce は、DataNode のブロックを「Map」(欲しいデータを取り出して分類する)と「Reduce」(分類されたデータを集計する)の2つのフェーズに分解して実行する (図 2)。

Hadoopでは、データを保持しているノード上で計算を行うことでデータ転送量を抑え、同じノードで計算できなくても、一番近いノードからデータを取得することで通信を最適化し、高速処理を実現している。このように Hadoop によって分散



処理フレームワークが提供されるため、プログラマは Map 関数と Reduce 関数のロジックに集中でき、開発効率を高められるというメリットもある。

以上の HDFS や MapReduce に加え、データ管理 / 操作用語の「Hive」や「Pig」、NoSQL データベース (後述) の「HBase」など、様々なサブプロジェクトが立ち上がっている。また、同じオープンソースの Linux 同様、商用のディストリビューション (配布パッケージ) が多数登場しており、一般企業の導入環境も整備されてきている。

### RDB の課題を解消しビッグデータに最適化された「NoSQL」

ICT システムの多くは、複雑な検索や集計を得意とするリレーショナルデータベース (RDB) を使用しているが、ビッグデータを扱う上では、非構造化データの処理や、スケールアウト面での課題があった。そこで Amazon や Facebook を始めとするネット企業は、自社サービスの目的に沿った新しい DB システムを開発するようになった。

RDB がデータ操作に SQL を使うのに対して、新しい DB システムは SQL を使用しないことから、「SQL に取って代わる」という意味合いで「NoSQL」と呼ばれる。しかし最近では、RDB が得意でない部分で補完的に使用されることから、「Not only SQL」と捉える見方もある。

NoSQL では、RDB が重視するデータの一貫性を犠牲にすることで処理を高速化しており、データ構造も非常に単純である。また、データ構造を最初から定義する必要がない点や、スケールアウトによる拡張性を実現できる点なども、RDB と異なる (図 3)。

	RDB	NoSQL
データタイプ	構造化データ	非構造化データがメイン
スキーマ	事前定義が必要で、固定的	事前定義不要で、柔軟に変更可能(スキーマレス)
データ一貫性	ACID 特性により一貫性が厳密に維持されている	一時的に一貫性が厳密に維持されていない状態がある(結果整合性)
拡張性	スケールアップが基本。データ一貫性の維持を厳密に行うため、パフォーマンス低下が目立つ	スケールアウトにより大量アクセスによるパフォーマンスの低下が少なく、リニアにスケールする
サーバ	1台で稼働することが前提	分散・協調して動くことが前提
耐障害性	耐障害性を高めるためのコストが高い	単一障害点がないものが多く、低コスト
問合せ言語	SQL	SQLでない複数言語サポート
データ量	大規模データの高速処理にはパーティショニングやINDEXを張るなどの工夫が必要	大規模データの高速処理が容易

図3 RDBとNoSQLの違い

NoSQLに分類されるDBのデータストア方式には、キーバリュ型(KVS = Key-Value Store)、カラム指向型、ドキュメント指向型、グラフ型がある(図4)。キーバリュ型(KVS)の1つ「Dynamo」は、Amazonのショッピングカートで採用されている。カラム指向型は、大量のデータを対象とした分析やデータマイニングに特に適しており、FacebookやTwitterなどで採用されている「Cassandra」や、Hadoopのサブプロジェクト「HBase」がある。

キーバリュ型(KVS)	キー バリュ	キーとバリュを一組とするシンプルな構造。スケールアウトによる性能拡張に向く
カラム指向型		1つのキーが複数のカラム(列)を持つ。データ構造の変化に柔軟に対応できる
ドキュメント指向型		JSONやXMLなどドキュメント形式でデータを管理。ブログや記事の管理などに向く
グラフ型		ノード、リレーションシップ、プロパティで構成。データ間の関係性を管理することに向く

図4 NoSQLの種類と特徴

## ビッグデータのリアルタイム分析を可能にする「CEP」

ビッグデータ処理には、Hadoopに代表されるバッチ処理の他に、時々刻々と流れてくるデータストリームをリアルタイムに分析する「ストリーム処理」もある。その技術として代表的なのが、「CEP(複合イベント処理 = Complex Event Processing)」である。

CEPは、複数の情報元から絶えず生成されるデータをハードディスクに格納せず、メモリに読み込む。そして「CEPエンジン」にあらかじめ設定しておいた「分析シナリオ」と条件が一致すれば、次のアクションへとつなげていく。たとえば、大手ショッピングモールでは在庫状況をCEPで監視し、在庫が多い日は夕方から特売セールを実施し、顧客の位置情報

と嗜好から近くにいるターゲットのモバイルにメールで告知することで、顧客の購買行動を喚起している(図5)。

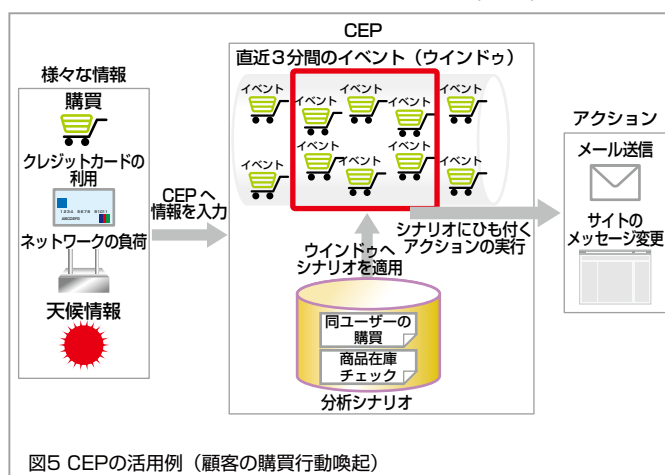


図5 CEPの活用例(顧客の購買行動喚起)

CEP自体は新しい技術ではないが、近年の新たな動きとして、Facebookの「Data Freeway」やTwitterの「Storm」、Yahoo!の「S4」など、ネット系企業が自社用に独自のCEPを開発し、オープンソースとして公開している点が注目される。

## データサイエンティストとデータ分析

ビッグデータ基盤やツールをそろえても、それ自体がビジネスの価値創造につながるわけではない。ツールを使いこなし、膨大なデータの山から隠された金脈を探し出し、ビジネスにまでつなげる人材が必要となる。こうしたスキルを持つ人材が「データサイエンティスト」と呼ばれており、GoogleやAmazon、Facebookなどが成功してきた背景には、データサイエンティストの存在がある。

データサイエンティストに求められるスキルは、統計解析やデータマイニング、機械学習といった「分析スキル」、分析ソフトウェアやHadoop、非構造化言語を扱うためのプログラミングといった「ICTスキル」、そして、自社のビジネスに必要なデータや分析を見極める「ビジネス知識」と、広範囲にわたる(求められるスキルや人材不足の状況、および育成事例などについては「巻頭言」参照)。

データ分析の流れは、課題を設定して必要なデータを特定・収集・加工し、データに潜むパターンや規則性を表すモデルを作成、そのモデルの検証結果からビジネス施策を立案していく、という流れとなる(図6)。

①課題設定	データ分析によって解決したいビジネスの課題を設定
②分析設計/データ設計	ビジネス課題を分析の課題に置き換え、必要なデータを特定する
③データの準備	分析に必要なデータを集め、分析可能な形に加工する
④モデルの作成	データを分析したり、ルールを見出したり、モデルを開発したりする
⑤モデルの検証	分析結果を解釈する。場合によっては、チューニングや再実験を繰り返す
⑥分析結果の活用	分析結果を元に、ビジネス施策を立案・実行する

図6 データ分析の流れ

※1 ACID 特性: Atomicity (原子性)、Consistency (一貫性)、Isolation (独立性)、Durability (耐久性) の頭文字をつなぎ合わせたもの  
 ※2: RDBでもカラム指向型を取り入れているものもある

富士通の取り組み

ビッグデータ活用を成功に導く

目的の明確化からお客様を支援

ビッグデータ活用において目的の明確化は重要なポイントとなる。目的を明確にすることで効率的なビッグデータ活用が可能となり、企業経営に新たな価値をもたらす。

富士通は、昨年開設した「ビッグデータイニシアティブセンター」を中心に、お客様のビッグデータ活用を支援しており、お客様の課題に最適な提案で応える活動を行っている。河合センター長に、企業におけるビッグデータ活用の現状についてお話を伺った。



河合 美香 氏

富士通株式会社  
統合商品戦略本部  
ビッグデータイニシアティブセンター  
センター長

ビッグデータ活用では  
目的の明確化が重要なポイント

システム導入において、「システムありき」ではなく、「システムを導入しどのように課題を解決するか」が重要であるように、ビッグデータ活用でも「データありき」ではなく、「データを活用しどのように経営に貢献するか」といった目的の明確化が重要なポイントとなる。ビッグデータ活用の目的は、高度な分析を行うことではなく、分析結果から潜在するパターンや法則を発見し、次のアクションにつなげ、企業経営に新たな価値をもたらすことにあるからだ。

また、調達、生産、物流、販売などの社内データに加え、センシングや位置情報、ソーシャル、気象、交通、行政など多種かつ膨大なデータから、目的や仮説もなく「新たな気づき」を見つけ出すのは至難の業だ。目的を明確にすることで、精度を高めるために必要なデータの選択や、最適な分析手法との組み合わせが可能となり効率的なビッグデータ活用が実現できる。

企業はこれまでBI (Business Intelligence) などを使ってデータを分析してきたが、ビッグデータ活用ではデータの規模やスピードが大きく異なる。例えば、月次レポートだったものがリアルタイムに把握できたり、24時間365日、刻々と変わるセンサーのデータが利用できるなど、今まで分析できなかったことが可能となる。また機械学習<sup>※1</sup>の進化により人間の思考では発見できなかった新たな気づきを得ることもできる。

企業はどのようにビッグデータを活用したいと考えているのだろうか。富士通の商談においてビッグデータ活用のテーマを分析した結果、最も件数の多いテーマがマーケティング、次に故障予測、新サービス創造、行動分析、需要予測と続く。企業のマーケティング部門や工場などからの問い合わせも増えるなど、現場での活用ニーズは高まっている。スマートフォンやタブレットなどモバイル端末の普及とクラウドの普及により、現場で情報を活用する環境が整ってきたことが背景にあると言えるだろう。

※1 機械学習：人工知能における研究課題の一つ、人間の学習能力と同様の機能をコンピュータで実現する技術

しかし、現場は業務に関して詳しいが、分析やICTの専門家ではない。現場がやりたいことと、それを実現するために、どんなデータをどのように収集するかといったICTや、分析手法を組み合わせることが必要となる。

富士通は、ビッグデータを活用したい現場と情報システム部門を対象に仮説立案ワークショップを無償で行っている（図1のSTEP1）。ブレインストーミングによりニーズを抽出し、目的の明確化を図り、それを実現するための目標や効果のKPIの設定、仮説立案を行う。STEP1の仮説立案ワークショップで目的などを設定してから、データの選定や分析手法を検討するSTEP2、お客様のサンプリングデータを使って分析し検証を行うSTEP3、実装フェーズに向けて企画構想や立案を実施するSTEP4へと、富士通はビッグデータ活用の検討段階からきめ細かくサポートしていくサービス商品も用意している。

ワークショップや実証検証は、ビッグデータイニシアティブセンターが主導で、富士通トラステッド・クラウド・スクエア内のビッグデータイニシアティブラボで行われる。



図1 ビッグデータ活用の目標・効果の設定から支援

課題解決メニューからプラットフォームまで  
体系化して提供

2013年6月、ビッグデータを活用したイノベーションを加速するために開設されたビッグデータイニシアティブセンターは、お客様、協業パートナー、ベンチャー企業向けの支援を総合的な

観点で行う。お客様支援では、コンサルタント（富士通総研）、要素技術・研究開発（富士通研究所）、データ分析の専門家キュレーター<sup>\*2</sup>、BIのSEやビッグデータプラットフォームの技術者など、富士通グループのビッグデータに関する技術やサービスを統合しお客様に最適な提案を行う役割を担っている。

検討段階の支援ではワークショップに加え、オフリングの活用も重要な取り組みとなる。約200件のモデル事例を通じて明確となった市場性の高い利用シーンを課題解決メニューとしてオフリング化し、富士通の商品やサービス、独自技術を組み合わせ実装モデルとして提供する。

具体的には、業務プロセス改革（リアルタイム経営の実現／故障予測による設備メンテナンスの高度化など）、サプライチェーンの最適化（需要予測の高度化によるサプライチェーンマネジメントの最適化）、マーケティングの高度化（顧客需要分析による人的リソースの最適配置など）、商品・サービス強化（M2Mデータによる商品・サービスの高度化）の4テーマ、10種のメニューとなっている（図2のオフリング）。

お客様の間では、「ビッグデータは活用したいが、どこから手をつけていいのかわからない」といった声も非常に多い。ビッグデータ活用は企業の競争力やノウハウに直結しており、扱うデータも機密性の高い情報が多いため、企業が参考にした事例数が少ないのが現状だ。過去の商談をベースにしたオフリングメニューは自社でビッグデータを活用するイメージを抱きやすく、お客様のビッグデータ導入の糸口として有効である。

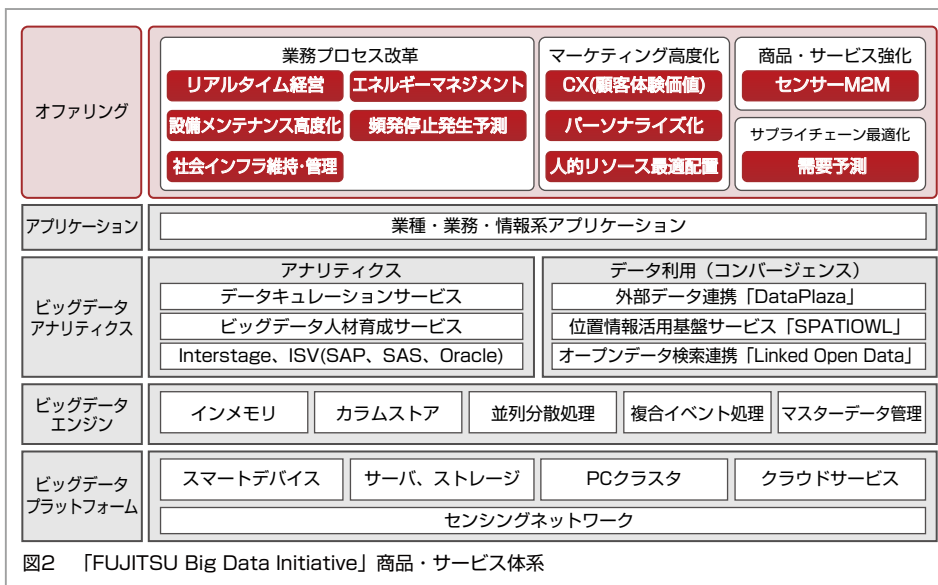
富士通は、オフリングからアプリケーション、アナリティクス、エンジン、プラットフォームまで、ビッグデータ関連の商品やサービスを「FUJITSU Big Data Initiative」として体系化し（図2）、お客様のビッグデータ活用の最適化を実現していく。

## データサイエンティストチームづくりも支援

ビッグデータ活用の可能性は先進のICT技術により大きく広がる。位置情報にSNSなどの情報を重ね合わせ、新たな価値を創出する位置情報活用基盤「SPATIOWL（スペーシオウル）」、世界中のデータ公開サイトのLOD（Linked Open Data）を格納し高速な一括検索を提供するLOD活用基盤、音響データによる故障予知といった最先端音響技術、収集から活用まで様々な要素技術など、富士通はビッグデータ関連の技術開発に力を注いでいる。

また、ビッグデータを活用したイノベーションは社会全体を巻き込んで拡大していくため、各業界をリードする企業や団体とのパートナー連携強化を図っている。さらに、ビッグデータを活用した新たなサービスを開発したいベンチャー企業を募集しビジネス化の支援も開始した。

ビッグデータ時代における競争力の強化では、データサイエンティストの育成も必要となる。しかし、業務、分析技術、ICTのすべてを理解するデータサイエンティストの育成は容易ではない。また、成果につなげていくためには組織として機能することも大切だ。



富士通では、自社内でも実践しているデータサイエンティストチームづくりを提案している。チームは、ビジネス企画担当、データ分析担当、システム担当によって構成されることから、既存の人材を活用し迅速に対応できる利点は大きい。データサイエンティストチームづくりを支援するビッグデータ利活用研修も定期的で開催している。

今後、富士通はビッグデータ活用により日本企業の競争力を高めるべく、技術・研究開発はもとより、オフリングの拡充などビッグデータイニシアティブセンターを中心にお客様支援を一層強化していく。

<sup>\*</sup>2 キュレーター：

従来の意味は、博物館や美術館等における学芸員のことを言う。富士通では、統計、機械学習、ICTスキルをもつ新たなデータ活用の専門家を指す

<sup>\*</sup>3 オフリング：

申し出、提供の意味がある offering。富士通ではお客様のニーズに応じた機能・パッケージ・機器を融合したサービスを指す。実績に基づき開発された具体的な課題や成果を明示している

<sup>\*</sup>4 データサイエンティスト：

データ活用の専門家。統計解析や機械学習、分散処理技術などを用いてデータを分析し、企業の課題を的確に把握し、解決策を提案する。本誌 10 ページ「技術解説」参照

### 関連サイト

●富士通のビッグデータ ～ Big Data Initiative ～  
<http://jp.fujitsu.com/solutions/bigdata/>

### ●実践事例に学ぶ!ビッグデータ利活用研修

日時：3月17日(月)～20日(木) [3月開催]

場所：富士通ラーニングメディア 品川ラーニングセンター

<http://jp.fujitsu.com/solutions/bigdata/seminar/20140114/>