

第4のメディアが作り出す ビッグデータの時代



東京大学教授
東京大学地球観測データ統融合連携研究機構長
東大生研戦略情報融合国際研究センター長
文部科学省科学官
工学博士

喜連川 優 氏
Dr. Masaru Kitsuregawa

【プロフィール】

1983年、東京大学工学系研究科情報工学専攻博士課程修了後、東京大学生産技術研究所に入所。2003年、生産技術研究所 戦略情報融合国際研究センター長に就任。ますます膨れ上がるデータの管理、情報洪水への対策に関する研究を推進している。現在、最先端研究開発支援プログラムを推進中。日本 IBM 科学賞（1991）、関係データ・モデルの生みの親である IBM E.F. Codd 氏の名を冠した ACM SIGMOD E.F.Codd Innovation Award をアジアで初めて受賞（2009）。

Professor
Director of Center for Information Fusion
Institute of Industrial Science,
The University of Tokyo
Executive Director, Earth Observation Data
Integration & Fusion Research Initiative
(EDITORIA), The University of Tokyo
Science Advisor, The Ministry of Education,
Culture, Sports, Science & Technology

ソーシャル・メディアへの投稿やインターネット上に保存される写真や動画などのさまざまな情報が爆発的に増加している昨今、膨大なデータを利活用することが、ビジネスの成功率と加速率に寄与する時代になっています。本記事では、以前より「情報爆発」を予見し、ビッグデータ関連の研究を重ねられてきた東京大学教授喜連川 優氏のインタビューを通じて、ビッグデータの有効性、活用方法あるいは活用や研究を推進するに当たった課題などについて取り上げました。

喜連川氏は情報爆発の大きな要因として、センサーなどを通じて取り込む情報を指摘。これを第4のメディアとして定義付け、その情報の分析からいかに高い価値を生み出すことができるかということがビッグデータのソリューションを考える上で重要であると力説します。

第4のメディアの「知」を基に新たなソリューションやサステナブルな社会システムをデザインしていくことは、IBMの提唱する Smarter Planet の世界につながっていきます。

IT が生み出した情報爆発の問題から その活用へ

— これまで文部科学省の「情報爆発時代に向けた新しい IT 基盤技術の研究（以下、info-plosion）」や経済産業省の「情報大航海プロジェクト」などに携わってこられました。この分野の研究を始められた経緯とこれまでの研究概要についてお聞かせください。

喜連川氏：21世紀に入ってから、顕著なスピードで情報量が増大するようになりました。この情報爆発ともいえる現象は IT 領域における近年最も重要かつ最大の変化の1つであると考えました。一方、この現象そのものを生み出してきた主役が IT であるのも事実であり、この状況に対して何らかの手を打つこともまた IT を担う者の責任ではないのかと考えたことが、この分野の研究を始めたきっかけです。2004年に情報爆発をテーマとした研究を文部科学省 科学研究費 特定領域研究として申請し、翌年に info-plosion が発足しました。

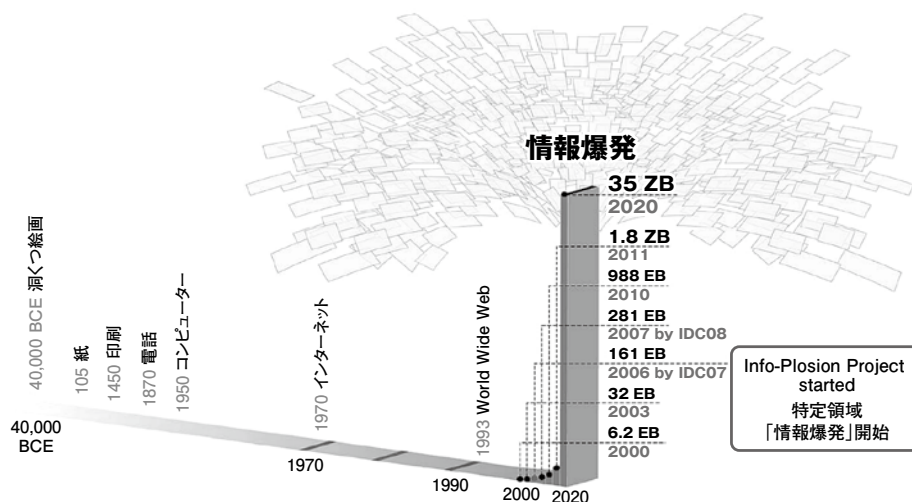
情報爆発という考え方については、わたしどもが研究を始める以前からもあったと思いますが、大きなプロジェクトとして、本格的に取り組んだのはわれわれが世界で最初だったのではないかと思います。この研究は、何か特定

の研究・開発目標を定めてそれに向けて一直線に進むというタイプのスタイルではなく、情報爆発という現象を取り上げ、それぞれの研究者がその現象に対して、どのような課題を解決する必要があるか自ら考え、あるいは、今後どのような問題が発生し得るかという点について多様な視点から議論を深め、1つ1つ問題を解き明かしつつ世の中に提示するというスタイルのプロジェクトでした。このようなスタイルの情報系のプロジェクトはわたしたちの知る限りにおいて過去にはなく、非常に大きなチャレンジとなりました。

info-plosionの取り組みは2005年に準備研究が始まり、2006年から300名規模の研究者で本格的にスタートしました。翌年には経済産業省の情報大航海プロジェクトが発足しましたが、これはinfo-plosionが始まったことによる影響もあったと思います。当初2つのプロジェクトが始動したころは、世の中では、情報爆発は人類にとって情報量がオーバーロード、いわゆる過負荷になるというネガティブな側面がクローズアップされており、爆発的に増加する情報をどのように管理すれば人が情報におぼれずに対応できるかという視点での議論が中心でした。しかし、逆に考えますと、人々がこれほどまでに膨大な情報を手にすることは人類史上初めてでもあり、それをチャンスと考えられるのではないかと。すなわち、オーバーロードというネガティブな側面だけにフォーカスするのではなく、膨大な情報だからこそ、情報爆発時代だからこそ生み出し得るまったく新しいソリューションも考えようという気持ちで両プロジェクトを推進してまいりました。最近では後者の視点の方がより重要視されるようになってきているように感じます。今では情報爆発を「ビッグデータ」と表現するようになっていますが、これは膨大なデータをポジティブにとらえる風潮が強まったことの表れなのでしょう。

第4のメディアに着目し 有効に活用することが重要

— これまで研究を始められたころに比べて、情報増大



出典: Horison Information Strategies, cited from Storage New Game New Rules, p.34 (www.horison.com), IDC, The Diverse and Exploding Digital Universe 2020 (<http://www.emc.com/collateral/demos/microsites/idc-digital-universe/view.htm>)

図1. 情報爆発

傾向にはどのような変化がありますでしょうか。

喜連川氏: 情報爆発といえば、SNS（ソーシャル・ネットワークワーキング・サービス）などの普及により、Web関連の情報の急増がその中心であろうとイメージされますが、この現象は実はそれほど大きな問題ではありません。人口の増加やコンピューターの普及には限界があるので、SNSなどの情報増加もそれに応じてピークを迎えるという指摘もあります。情報の爆発的な増加をもたらすドライバーは、センサーやRFIDから膨大なデータが次々と送られてくることです（図1）。

これをわたしは第4のメディアと呼んでいます。これまでコンピューター・サイエンスが過去30年、40年と研究の対象としてきたメディアは、人が目から取り込む情報、つまり画像や映像によるメディア、耳から取り入れる音声のメディア、そして文字として表記するテキストのメディアの3種類がありました。これらに対して非常に多くの技術が蓄積されてきました。

しかし、近年のセンサー技術のさらなる進展とネットワーク技術との融合により、「モノ」から送られてくる情報が新しいメディアとしての機能を有するようになったと見なせます。言い換えれば、「モノ」がしゃべるようになり、場合によっては人が発する言語以上に情報爆発を加速させ、重要になりつつあるといえるでしょう。これは、新しいシステムを考える上で非常に重要なポイントになると感じています。もちろんソーシャル・メディアも重要ですが、加えて「モノ」がしゃべるメディアをうまく活用することにより、新しいチャンスが生まれると考えています。

「とにかくためる」： 情報爆発を新たな価値へ

—— センサーなどからの情報が重要と考えると、それではその情報をどこまで保存しておけばいいのかという疑問が起こると思うのですが、その点についてはどのようにお考えでしょうか。

喜連川氏：これまでのデータベース関連の技術は、より多くの情報量を扱うことへのチャレンジを原動力に進化してきましたが、その技術を活用して、より多くの情報を保存し、後にデータ・ウェアハウスなどを活用することによってアドホックに価値を搾り出すべく有効に活用するという考え方は、米国においては一般的になってきています。例えば、世界金融危機以降も高い売り上げを維持している米国のECサイトなどでは、保持している膨大なデータを解析し、マーケット・ニーズがどこにあるのかということを常に把握することによって、ビジネスを成功させています。こうした分析結果は、ビッグデータを基に詳細な動きを綿密に解析することによって初めて明らかになるものですが、このように詳細なデータを保存しておくことが重要であるという認識は日本においても少しずつ広がりつつあるものではないでしょうか。「企業にとっての価値あるアセットは、ITシステムではなく、保存されている情報そのものにあり、ITシステムはアセットである情報をいかに活用するかというツールである」という考え方は次第に醸成されてきていると感じています。

そこで問題になるのは、どこまでの粒度で情報を保存しておけばいいのかということです。日々生成される情報量は、ストレージの容量をはるかに上回っていますので、すべてを保存することは原則不可能です。こうした状況の中で、収集したどの情報をどの段階で破棄して情報ライフサイクル管理（Information Lifecycle Management：ILM）を最適化するのが今後の大きな技術課題だと思います。最終的な取捨選択の判断は人の鋭いセンスに依存するところが大きいのも事実です。

ビッグデータとクラウドにより デザインされる新しい世界

—— センサーなどから取り入れる大量のデータを第4のメディアと表現されていましたが、メディアと呼ぶからには、それらの情報が人に何らかの影響を与えるような仕掛けが必要になると思うのですが、その点についてはどのようにお考えでしょうか。

喜連川氏：今日までのコンピューターの役割を振り返ると、開発された当初は高速に数値計算する機能が重宝され、その後はビジネス情報を効率良く管理するデータベース機能が社会に大きく貢献してきました。すなわち、コンピューターは人間には到底できない計算能力や記憶能力を発揮することにより価値が見いだされてきました。この時代のコンピューターの役割は非常に分かりやすかったと思います。しかし21世紀になると、メディアとしての役割が最も重要になってきました。これは、ほかの役割の重要性が下がったのではなく、メディアとしての重要性が相対的に高まったと考えられます。ブログ、Twitter、Facebookなど、さまざまなメディアが形成されてきた流れを考えますと、今後さらに多くのメディアが登場してくると予見されます。そして、それらのメディアの中にセンサーなどを通じて「モノ」から発信された情報が含まれるようになると、それは人に対して影響をもたらすと考えられるでしょう。

このようにコンピューターの役割の変遷を振り返ると、第4のメディアから取り込まれる情報をはじめとした膨大な

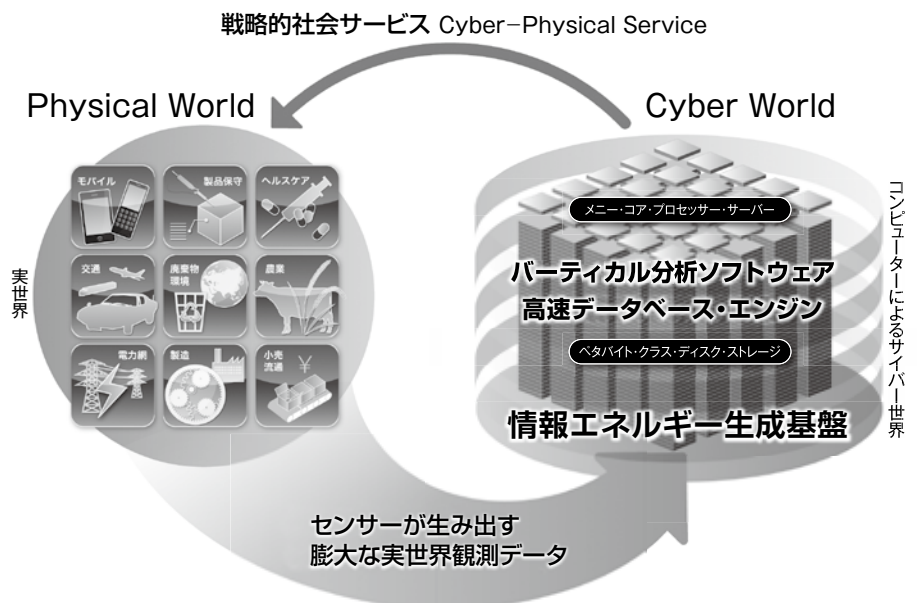


図2. 戦略的社會サービス「Cyber-Physical Service」

データを扱う技術は、今後の IT の根幹を支えていくのではないかと推察されます（図 2）。

またビッグデータの活用には、クラウド・コンピューティング（以下、クラウド）が真価を発揮するでしょう。クラウドはコンソリデーション（資源の集約化）による運用コストの低減がクローズアップされてきましたが、今後の IT はコストを下げるだけでなく、いかに価値を創出するかということにシフトしていくと思います。

クラウド環境を活用する場合、旧来のシステムからクラウド環境にいかにかデータを移行するのかということが常に問題視されます。確かに今ある膨大なデータをクラウドに移行する場合にはその手間からうんざりします。しかし、少し視点を変え、新しいビジネスをクラウド上で展開することを考えてみましょう。例えば 1PB（ペタバイト）のビッグデータを 1 年 365 日かけて蓄積すると想定すると、実際には 1 分当たりですごく少ないデータ量の収集にすぎないことが分かります。第 4 メディアを活用する新しい IT ソリューションを考えると、むしろ自然に 1PB 程度のビッグデータを構築することができるのではないのでしょうか。そして膨大なデータの蓄積という観点では、必要に応じてシステム規模を自在に拡張できるクラウドの強みが生かされるようになります。このように単なるコスト抑制としてのクラウドだけではなく、実世界を対象とした新しいビジネスを生み出す手段としてのクラウドが真価を発揮するようになるでしょう。巨大化を続ける蓄積データに対する多様な分析技術の精緻化とともに、さらなる新たなソリューションやサステイナブルな社会システムをデザインしていくフェーズへと移行すると期待しています。すなわち、分析とデザインが循環する構造を持つ世界が開けると思います。

エネルギーやヘルスケアの分野で始まった 第 4 のメディアの「知」の活用

— それでは、ビッグデータが新しい価値を生み出す具体例をご紹介します。

喜連川氏：1 つは東京大学 生産技術研究所の野城教授、馬郡研究員らが取り組んでいる事例を取り上げてみたいと思います。これはコンビニエンスストアと共同で研究し、店舗の省エネを実現するセンサー・ソリューションです。2009 年から施行されている改正省エネ法では、大きなビルだけではなく、多数の小さな店舗からなる事業者も対象となり、コンビニエンスストアなどが対象となることか

ら研究が着手されました。コンビニエンスストアの店員は、接客や清掃、売り場の管理、発注業務など常に何らかの作業をしている状況にあり、省エネに注意を払うことは事実上難しくなっています。また、コンビニエンスストアが設置されている環境は店舗ごとに異なるので、エネルギーの使われ方も店舗によって変わってきます。そこでセンサーを活用して店舗の電力使用状況をモニタリング・解析することにより、エネルギー削減ポイントを特定し、細やかな制御を行い、全体で年間を通じて 10% 以上のエネルギー削減効果が得られることが確認されました。このソリューションにより、店舗ごとのセンサー・データがクラウドに集積され、種々の解析が可能となります。知の蓄積により多様な店舗条件に対し最適な省エネが実現できる可能性が広がります。すなわち、新店舗設置の際の設計にも生かされると想定されます。そして新しい店舗からさらに情報が蓄積されるというサイクルが回れば、よりエネルギー削減効率が高まっていくことでしょう。この事例から分かるように、第 4 のメディアの「知」をいかにストックし、活用していくということがビジネスの成功率と加速率に大きく寄与するという時代にすでになっているといえます。

上述の事例では機器の入れ替えは行いませんが、新たに建物を設計する場合には効果ははるかに絶大です。現在、東京大学では完全なゼロ・エミッションを目指した教育棟を構築し、統合化最適制御技術も導入しつつ、運用を開始しています。すでに従来の平均的な教育施設と比較して 50 ~ 70% の運用エネルギーを削減し、今後さらに ZERO を目指しています。

ほかの事例として、医療分野のホームケア（自宅医療）での取り組みをご紹介します。ホームケアはさまざまな国で重視されており、例えば米国では、コスト削減の目標として施設医療（病院における医療）の 4 分の 1 をホームケアに切り替えることが掲げられているなど、各国でいろいろな試みがなされています。ホームケアでは医師や看護師が常駐することはできませんので、施設医療から自宅医療に移行していくということは、医療の世界におけるセンサーの重要性が増していくことを意味しています。

情報大航海プロジェクトにおいても、センサーを活用したホームケアの研究を行いました。実証内容としては、血糖値をモニターするセンサー、運動量を測定する加速度センサーを用いて、糖尿病患者に継続的に自宅で血糖値を計測してもらい、その値に応じて、「運動をしましょう」「食事量を抑えましょう」といった行動を促すメッ

セージ、これを「情報薬」と呼んでいます。情報薬を適切なタイミングで提供するというものでした。被験者の同意を取るなど種々の準備が必要となることから、実証期間はクリスマスや正月などのイベントが多い年末年始にかかることとなり、糖尿病患者にとっては血糖値が上がりやすい期間なのですが、情報薬を提供したこの時期においては、被験者の血糖値の上昇を完全に抑制することに成功し、行動を促す情報を適切なタイミングで与えることが薬と同等の有効性を発揮することを実証することができました。当時3年間の情報大航海プロジェクトと並行して5年間のinfo-plosionプロジェクトが進行していましたが、2つのプロジェクトを連携させることにより、この実証実験を行ったまったく同じ被験者の前後1年間ずつ計3年間のデータを取り扱うことができました。これにより、長期の経過観察が可能となり、糖尿病における情報薬の有効性をより明確にできたことは大変有意義でした。そのほかにも多くの実験をし、運動を促す情報薬によりほとんどの被験者で運動量が増えることなども分かりました。

この加速度センサーを常時用いた情報薬のアプローチを、例えば全国約2,000万人のメタボリック症候群の方に適用しようとする、膨大なセンサー・データ（ビッグデータ）を処理する必要があります。しかし、皆が自分自身で運動量をチェックし、自身の健康をコントロールできるようになる時代が来ることもそれほど遠い夢ではないことを感じさせられる実験ができたと思っています。これは従来では考えられないような規模のビッグデータとその基盤を用いた、医師と患者とITを連動した新しい形態のヘルステックともいえるかと思っています。こうした時代感覚を医療従事者、患者、そしてIT関係者という3者のステークホルダーで共有していくことが今後ますます重要になると考えています。

ビッグデータによる新しい世界に向けて、 人材育成、プラットフォームの構築、 法整備が重要

—— ビッグデータの基盤作りという話題が上がっていましたが、さまざまな分野においてビッグデータを効果的に活用するためには、人材育成をいかに進め、またどのような形でプラットフォームが整備されていることが望ましいとお考えでしょうか。

喜連川氏：ビッグデータにかかわらず、あらゆる分野において、問題は最終的には人材育成に帰着するといっても過言ではありません。ビッグデータについても、人材に関する洞察を深めていく必要があるでしょう。そこで重要なポイントは、ビッグデータを解析できる技術者と、その分野の専門家という双方のステークホルダーが協業できる場をいかに作り出すのかということだと思います。

例えば、さまざまな分析ツールを駆使できる技術者は、その分析を行う対象について必ずしも詳しい知識を持っているわけではありません。スマート・シティであれば、水、交通、大気汚染、物流、エネルギー、犯罪、健康など、それぞれにおいてステークホルダーがいますが、技術者がそのすべての分野の知識を十分に身に付けることは到底不可能です。

このことを踏まえると、それぞれの分野の専門家が、ITスキルおよび分析のリテラシーをある程度向上させるということが有効になってくるのではないかと考えています。そして分析の最先端となる部分については分析技術の専門家に依頼するという形で協業することが比較的妥当な方法だと思います。つまり、人材育成はIT側だけで考えるのではなく、その活用対象となる領域の専門家の方でも取り組んでいくことが効果的だと思います。

同様にプラットフォームについても難しい問題があります。企業の経営者にとって、利益を生むという確証が得られていない段階でビッグデータに投資し取り組むということは、非常に大きなリスクを伴います。現在の日本の経済状況を考えますとさらに難しいといえるでしょう。

こうした状況の中で、企業や、大学をはじめとした各種研究機関、文部科学省や経済産業省などの国の機関が、ビッグデータの手応えを体感できる実験の場が用意されれば、ビッグデータの活用促進を後押しすることになるのではないかと思います。可能であれば、国などの機関がそのような取り組みを後押ししてもらえれば、効果的ではないかと考えています。

さらにデータの権利が誰に帰属しているかということが明確になっていないという問題もあります。センサーを設置した人なのか、センサーを設置した場所の所有者なのか、あるいはセンサーが設置された場所でビジネスを行っている人なのか、さまざまな解釈が成り立ち、いずれも否定できない状況にあります。しかし、明確でないからといって放置しておいてよいことではありません。法が整備されていないことが原因で新しいビジネスの検討に消極的になり、日本でもこれまでに大きな失敗をした経験があります。

例えば、サーチ・エンジンの開発において、著作権法上、そのサービスがグレーであることが委縮効果を生み、日本が大きく遅れを取らざるを得なかった苦い経験があります。

また、ビッグデータといっても、一種類のデータだけを対象とした活用もありますが、異なる種類のビッグデータを融合する（掛け合わせる）ことにより、一層、わくわくする真に新しい価値が生み出されることが期待されます。業
際ビッグデータともいべきものでしょうか。しかし、企業と企業のデータの融合は情報の第三者への提供と見なされる可能性があり、同様の問題が生じます。

これらのデータの取り扱いに関してはさまざまな課題があり、慎重に議論しなければなりません、少なくとも国民が合意できる部分については、段階的にであっても積極的に法整備を進め、分かりやすいガイドラインを策定していく必要があります。これを推進しないと、法整備の遅れが日本経済の活力を阻害する大きな要因になってしまう恐れがあります。つまりITの進化のスピードに対して、法が緩やかにしか整備されていないという現状を正視する必要があります。もちろん法整備にすべてを押し付けてはいけません。同時に技術のアクセプタビリティ（受容性）についてIT技術者は当然のことながらもっと努力を払う必要があります。ITソリューションを開発する場合、それが世の中に与えるインパクトや、それを取り巻く法制度への影響を検証し、もし影響があるのであれば、ソリューションの有効性に対する理解を促していくことに惜しみなく努力を払わなくてはならない時代になっているといえます。

このように法整備とプラットフォームの構築の2つをパッケージ化して取り組んでいけば、ビッグデータの分野は非常に早いスピードで発展していくと思います。逆に、それがうまくいかずにビッグデータの活用がなかなか促進されない危険性を非常に危惧しています。

ビッグデータの世界がもたらす 新たな技術のチャレンジ

— IBM では、IBM InfoSphere BigInsights や IBM InfoSphere Streams をはじめとしたビッグデータ関連の数々の製品や技術を開発していますが、ビッグデータ時代の技術、そして、ビッグデータ時代におけるIBMをどうぞ覧になられていますか。

喜連川氏：ビッグデータを活用するための技術はさまざまなものがありますが、それらを評価する前提としては、ビッ

グデータがもたらす価値を把握することが重要になります。また、大切なのはデータなのか、あるいはアルゴリズムなのかという点も重要なポイントです。少量のデータを対象にアルゴリズムを精緻化する競争が過去長らくなされてきました。しかし最近では、対象データ量を大きくすることによる分析の精度の向上がアルゴリズムの工夫をはるかにしのぐ事例が多く出てきています。逆にいえば、ビッグデータを処理するアルゴリズムはそれほど複雑でなくても、新しい発見につながるケースがあるということが出来ます。

ここで、情報爆発の研究を通じて経験してきた、テキスト分析に関する興味深い事例を1つご紹介したいと思います。情報爆発プロジェクト以前にテキスト分析を行う場合は、新聞社から過去すべての新聞を購入して得た2,000万程度の文章を対象とすることが通例でした。しかし、info-plosionプロジェクトの研究では、もちろんその当時はビッグデータという言葉などはなかったわけですが、Webサイトのテキスト約150億の文章を分析の対象としたのです。これだけ対象とする文章を増やしたところ、日本語分析の精度が飛躍的に向上しました。この結果によりビッグデータを活用することの効果の大きさを実証することができたのですが、その一方で、単にデータ量を増やしても分析性能が上がらないケースも判明しました。新たな課題を把握することで、次の技術開発に向けたチャレンジが生まれています。

IBMは、これまでのITの進化の中で常に新しい技術をけん引する役割を果たしてきた企業だと思っています。わたしがデータベースに興味を抱き研究を始めた学生時代は、日本には必ずしも十分な教科書がありませんでした。当時六本木にありました日本IBMのライブラリーによく通ったことを思い出します。われわれの研究分野において、最も積極的に国際的な学会の場で論文発表を行った企業はIBMであることは間違いありません。その意味で、IBMはとてもシンパシーを感じる会社です。また、「RISC (Reduced Instruction Set Computer)」や「RAID (Redundant Array of Independent Disks)」など、さまざまな新しいコンセプトもその原点をたどっていきますとIBMの研究にたどり着くことが多いという点でもその底力が感じられます。ビッグデータやSmarter Planetにおいても、Watsonのようなサプライズを見せていただけるのではないかと強く期待しています。