



知識社会を構築するビッグデータと政策のための科学

(独) 科学技術振興機構 社会技術研究開発センター シニアフェロー
奥和田 久美

1. はじめに

日本の人口動態変化は急速で、日本国内の市場性において、また国内の雇用情勢において、企業の戦略を大きく変化させるものになっていることを前号にて議論した(本誌9月号参照)。我々が強く認識しなければならないことは、過去の延長にはない世界がやってきたということであり、我々が有している過去数十年程度の個人的経験だけで対処できないことが増えているという現実である¹⁾。では、これから世界はどういう方向へ向かっているのだろうか？

マネジメント研究で今なお第一人者として人気の高いP.F. ドラッカー氏の最後の著作「ネクスト・ソサエティ 歴史が見たことのない未来がはじまる」²⁾は、すでに10年以上前に発表されたものであるが、これから社会に起こりつつあることを的確に予見しているように思える。彼によればネクスト・ソサエティとは知識社会であり、知識こそが資源となる。特に日本にとって、まず必要なイノベーションは「社会的な意味での革新」であり、知識労働者すなわち高等教育を受けた人々をいかに生産的な存在にするかが鍵となり、それが日本経済のゆくえを決めると言う。

本連載ではすでに、「科学技術イノベーション政策のための科学」において、種々のデータの蓄積と公開の進展が大きいことを紹介した(本誌6月号参照)。このようなデータの集合は社会全体が共有できる形式知であるが、将来的

に状況がさらに進展すると、これらに加えて、個人個人が有している経験・技能・暗黙知を含めた広い範囲の潜在的な知識が顕在化され、それらが我々を知識社会へと導くと考えられている。このような移行の兆候を絶妙に言い当てることで注目されているキーワードが「ビッグデータ」である。ビッグデータの恩恵はすでに我々の暮らしに現れはじめ、本誌でもこのキーワードが頻繁に登場するようになった(本誌8月号、9月号参照)。

ビッグデータには産業界の方々もかなり関心を寄せていると思われるが、懸念すべきはICT業界や情報産業のなかだけの話題、あるいは統計学や情報学のなかの技術テーマであると誤解した議論を見かけることである。ビッグデータの価値や影響は、我々の日々の暮らしの多くの場面に現れるものであり、多くの産業に関与し、おそらく第一次・第二次産業や公共サービスにこそ大きな変化をもたらす。また、それに先立って、多くの科学に革新的な知識がもたらされると予測されている。欧米では、これらの動きを第4の科学および第4の産業革命の到来、と見なすような国も出てきている。

本号では、急速に注目が高まるビッグデータにおいて、特にその「価値創出」の観点に注目し、世界で起こりつつある変化と「科学技術イノベーション政策のための科学」の研究開発プロジェクトとの関係性を議論していく。なお、ビッグデータに関連する要素技術や産業分野ごとの新ビジネスの可能性については、すでに多くの

著作や参考文献があり、それらも参考にしていただきたい。

2. 産業界で急速に高まった ビッグデータへの注目

本誌6月号では、オープンデータポリシーとともに各国でデータの公開や整備が急速に進みつつあることを紹介した（本誌6月号参照）。更新が適切に行われるならば、データは情報量の増大に応じて、徐々にその量や範囲を拡大していく。量的に大きなデータが必ずしも重要なデータであるとは言えないが、量的に大きなデータがこれまでに明らかでなかった意味合いや価値をもたらす例が数多く出てきており、既存ビジネスの助けになるだけでなく、それ自体が新ビジネスにも発展している。

ビッグデータの特徴は、Volume（これまでとは桁違いの多量性）、Velocity（発生頻度・更新頻度およびリアルタイム性）、Variety（これまでの概念を超える多種性）の「3V」にあると言われており、この3Vから生まれる新たな価値や革命的变化の可能性が、ビッグデータの注目される理由である³⁾。ここで、データの正確性や完全性がビッグデータの特徴ではないことは注意を要する。3Vがデータセットの不完全性や一部のデータの不正確性を補い、それら以上の価値を生むのである。また、この3Vを必ずしもすべて備えていなければいけないわけではなく、もっと広い意味でデータの公開や利用が進む動向をビッグデータと捉えている場合もある。ビッグデータというのは一種の流行語のようなものと見なす人もいるが、このような言葉を用いるかどうかはともかく、その意味するところを看過するならば、データから見出される新たな価値の可能性を見失う。

このような言葉に急速に注目が集まった背景には、もちろん、蓄積情報量の爆発的拡大がある。有史以来、各種の技術の発展により蓄積情報量の拡大傾向は一貫して続いてきており、特に紀元前後の紙の登場や15世紀の印刷技術の登場は、人類の知識の拡散に対して革命的に貢献したと言える。しかし、なんと言っても1990年代後半のインターネットの普及によ

る蓄積情報量拡大のスピードは桁違いであり、2010年代にウェブ上に蓄積された情報量はゼタバイト（10の21乗、2の70乗）のレベルに達している。2000年以降の蓄積情報量の増大傾向については、まさに「爆発的」という表現がふさわしい。もちろん、情報技術の進歩がこの増大を支えている基盤であり、特にその処理と通信の速度向上、そしてなによりもコスト低減効果の貢献が大きい。例えば現在、私たち誰もが手にしているスマートフォンの情報処理能力は、1980年頃のスーパーコンピュータの20倍以上である。記録コストは至近の10年間だけでも2桁は低減し、現在はギガバイトあたり1円程度になってきている。このような低コスト化が図られた結果、すでにビジネスとしては、情報処理能力や記録密度の向上の面だけから大きな利益を上げることが難しくなりつつある。また、通信速度向上は情報蓄積の仮想化・クラウド化を進展させ、データの個別保存や個人的記録所有の必要性を消失させつつある。SiriやWolfram Alphaの利用は、好きな時に最速のスーパーコンピュータを対話という形で使用していることになる。また、ネットワーク化されたテキストや画像を含む膨大な蓄積データは、それ自体が人工知能の進展をさらに加速している。

情報のオープン化や記録の無料化が進む一方で、無料ではないまでも購入可能なデータも増えていることを考えると、データ所有の既得権というものは徐々に失われている。もはや、データを所有すること自体は、セキュリティコストを生じるマイナス要因にすらなると認識せざるを得ない。企業にとっては、このコストに見合う以上の価値を創出できないならば、データを所有する意味がない。

今後起こり得るデータ内容の変化については後述するが、ビッグデータの価値の議論で無視してはならないのは、将来的には、これまで人に帰属すると考えられていたノウハウや暗黙知が形式知化していくこと、無意識の人の行動や自然界の隠れた法則性などが定量化・可視化されていくことによって、知識産業や知識社会が構成されていくことである。情報がデータ化・統合化・ネットワーク化されることによって、意識されていなかった多くの価値が生まれつつ

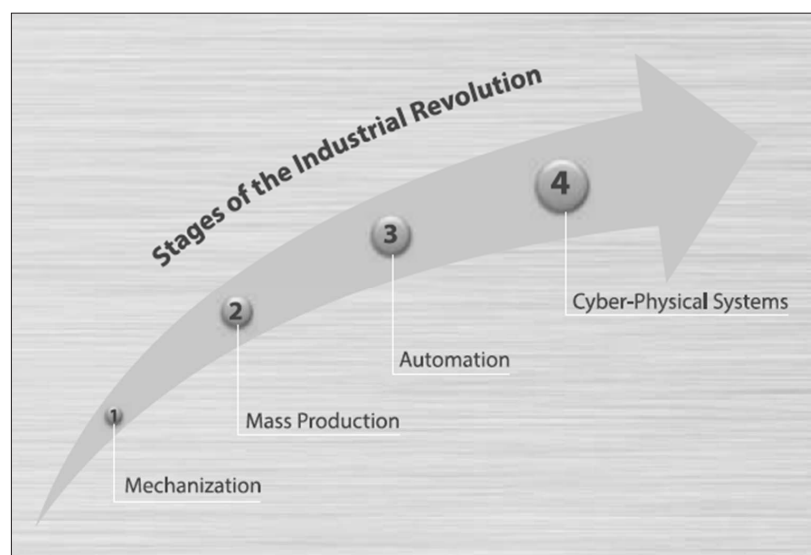


図1 Smart Factory が導く第4の産業革命⁴⁾

ある。

産業界における情報化社会の恩恵は、これまでは主に第三次産業、特にサービス産業や管理部門の業務などにおいて見られたが、最近、世界的に注目されているのは、これまで手付かずであった第一次産業における情報化の恩恵である。また、すでにかかなりの機械化・自動化が進んでいる第二次産業でも、さらに次の産業革命が誘発されるのではないかと議論になっている。言い換えれば、第一次・第二次産業においても知識産業化が進んでいくと考えられているのである。

例えば、日本と同じように製造業の強みを特徴とするドイツでは、「ハイテク戦略 2020 行動計画」の一環として、あらゆるデータをネットワーク化させることで製造を支える現場を「考える工場 (Smart Factory)」⁴⁾に変化させていこうという Industrie4.0 計画が始まっている。これは第4の産業革命を意識したものと言われている(図1)。

彼らによれば、生産されるモノ自体がデータを発生させ、生産プロセスを指示し得る要素となり、工場全体を「考える工場」に変化させることで、格段の生産性向上を図ることができる。これは、人間が関与しなくても生産・供給・製造パフォーマンスを最適化する Cyber-Physical Systems を構築し、先進国の最大の弱点である労働コストを大幅に減らすことを可能にするという考え方であり、技術

者や材料資源の不足を補い、資源効率の画期的向上によって、製造業の国際競争力を向上させる方法論と位置づけられている⁵⁾。当然ながら、これまでの職業を失う人も出るが、それに対応するための EU 域内の職業マッチングサービス提供プログラムも検討されている。生産性向上の視点は新技術の商品化速度の向上にも向いていることから、対象プロセスは製品の生産プロセスだけでなく、機能設計やデザイン、研究開発などの創造的プロセスにも発展し得る。一般的に、これまでは「何を作るか」と「どう作るか」が別次元のプロセスとして議論されているが、おそらく Industrie4.0 が進展した世界

ではこれらが有機的に繋がっているはずである。要するに、これはドイツの製造業を労働集約型から知識型産業に変化させるということを意味する。

この計画のなかの各技術要素は、一見すると第三次産業革命の自動化技術とあまり変わらないようにも感じられるのだが、上記のように、その意図するところはかなり異なっている。彼らの議論に照らし合わせるならば、残念ながら、日本の生産性向上技術や技術経営議論のほとんどは、まだ第三次産業革命の時代の議論が繰り返されているに過ぎないことになる。ドイツの計画が彼らの意図する方向へ進展するならば、このような考え方の存在は、日本とドイツの将来の製造業を大きく隔てることになるのかもしれない。

3. 暮らしが変わる

我々の日常生活では、ビッグデータからの恩恵を受ける機会が日増しに増え、新ビジネスが次々と生まれている³⁾。天気予報は、かつては気象庁や気象予報士だけが教えてくれるものであったが、現在では各地のごく一般の人からの情報提供だけでも日本全国の天気を知ることができる。また、今どこで風邪が流行りつつあるのかをつぶやき表示でリアルタイムに知ることができ、個人個人の予防行動にも風邪薬の出荷情報にも役立っている。また、特定の単語の検索が行われたことを統計的に追えば、インフル

エンザの流行分布が把握できることが判明し、地域的流行の予兆・ピーク・終息も把握できるようになった。現在は、自然災害情報・事故現場の様子が、メディア報道や政府公式発表よりも早く地図上に現われる。報道機関が到着するよりも前に、現場に居るごく普通の人々がその場の動画を即座にアップロードしてくれるようになったため、メディアのほうでそれらを利用させてもらうケースも多い。位置情報を伴って発信された人々の情報は公開され、後日、集合情報として統計的に地図上で解析されている。大地震発生が不可避の日本では、東日本大震災の際に発信された個々情報の統合データが、次の震災の備えとして準備されつつある。このように、ビッグデータは公共サービスへの影響力も大きい。日本の自治体や政府機関の電子化が遅れぎみであるだけに、特に人口減少の進む自治体の行政サービスでの利用価値が高いと考えられる。政府関係では透明性の確保という意味においても大きな寄与がある。

すでに十分に揃っているデータを統合するだけでも、かなり大きな社会的価値の創出が期待できる。例えばレセプト（医療報酬明細書）データをより広範囲で統合的に利用すれば、病院の効率的経営や医療費全体の無駄削減に繋がるのではないかと期待がある。日本では医療保険制度が普及しているだけあって、病院・調剤の99%、診療所の95%が電子化されており、フォーマットが標準化されたレセプトデータが揃っているそうである⁶⁾。このように品質の高いレセプトデータの揃った国は世界でも数少ないとのことで、知的財産としては長らく宝の持ち腐れとも言える状態にあった。すでにレセプトデータからの各種感染症流行の実態調査（サーベイランス）が始められている⁷⁾。また、電子カルテ・入院情報・救急車搬送・一般医薬品・保育園・学校・高齢者施設などの状況・各種検査などの各種情報源からの症状の早期探知サーベイランスも考えられている⁸⁾。政府の社会保障制度改革推進本部では、レセプトデータの統合がきっかけとなって、このままでは破綻しかねない社会保障制度の見直し議論につながることを期待している。

4. アカデミアにおけるビッグデータへの注目

ビッグデータは、ビジネスの世界だけでなく、もちろんアカデミアの世界にも影響を与えている。世界のアカデミアが注目したきっかけは2012年3月の米国政府による「ビッグデータ研究開発イニシアティブ」^{9) 10)}の発表ではなかっただろうか。これは、米国の喫緊の課題解決に役立てることを目的に、大規模で複雑なデジタルデータから知識や洞察を引き出す能力を高めるという研究開発の推進であった。このイニシアティブは、オバマ政権の科学技術イノベーション政策の5つの柱の1つであり、ここでは、ビッグデータはインターネットと同等のインパクトを世界にもたらし得るものと見なされた。具体的目標は、①大量なデータの収集・蓄積・保存・管理・分析・共有のために必要となる最先端の革新的技術を前進させる、②それらの技術を科学と工学における発見の速さの加速・国家安全保障の強化・教育と学習の変容のために利用する、③ビッグデータ技術の開発とその使用に必要とされる労働力を増強する、というもので、米国の各省庁やファインディング機関は、各研究分野において対応施策と支援方針を発表した。特に医学や生物などバイオ系の研究領域で上記③にあたる人材トレーニングが大規模に計画されたことが注目される。また、このイニシアティブはその後のEUの政策などにも影響を与えている。

エルゼビア社の学術文献データ Scopus を使ったビッグデータ関連の分析¹¹⁾によれば、タイトル・要約・キーワードに「big data」というワードを含有する文献は、2010年以前はほとんど存在しないが、2012年頃から急速に集合体を成すようになっており（図2）、上記のイニシアティブやその前後の議論の影響が感じられる。ただし、それらの参照文献の推移を見ると、1990年代から2010年までに徐々に蓄積されてきた学術文献が基になっており、情報科学だけでなく、ソーシャルネットワークなど社会とのつながりを持ち、医療・バイオ・脳研究などとも関連性を持って成長してきたことがわかる。また、共通性の高い中心的なワードにはマシンラーニングや知識発見といった言葉がある。

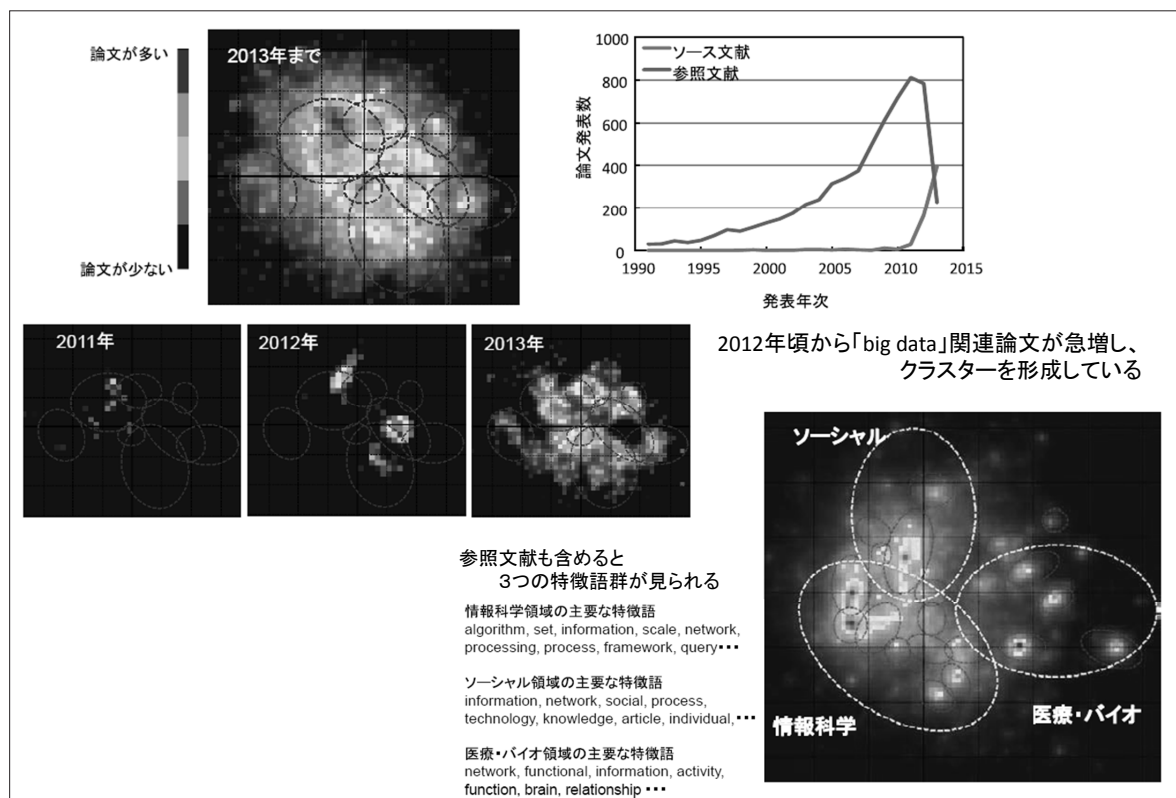


図2 エルゼビア Scopus によるビッグデータ関連文献の解析
(VALUENEX コンサルティング株式会社)¹¹⁾

ビッグデータという言葉は、アカデミアのほとんどにとっては突然現れた社会的流行語のようなものであるのだが、それ以前にすでにデータの多様性において拡大する発展トレンドが存在したわけで、そのトレンドをうまく表わし、かつ発展させるキーワードが現れたということなのかもしれない。

5. 科学が変わる・研究が変わる

情報のノーベル賞とも言われるチューリング賞受賞者のジム・グレイ氏は、1990年代に「ビッグデータのチャレンジは、科学の世界からやってくる」と述べていたそうである。マイクロソフトリサーチ社が2009年に編集した「第4のパラダイム」によれば、第一の科学とは観察、第二の科学とは解明、第三の科学とはシミュレーションであったが、起こりつつある第四の科学は「データ集約による科学的発見」であると言う¹²⁾。宇宙・地球・生命などありとあらゆる科学の領域で、膨大なデータの集約処理から、桁違いの知識が、しかもかなりのスピード感をもって得られるようになり、人類の知の世界がこれまでとは全く違うフェーズに入ってきてい

る。この結果、「研究」という行為自体が、ある一点の凝視から、線・面・空間・時間のような俯瞰性を重要視するメタ的あるいはシステムのなものに変容しつつあり、そのような研究活動からも新たな知的発見が生まれるようになっている。

まずは、研究のアウトプットとして認められる「研究の生産物」に多様性が増し、特許や論文のような結果を記した文献だけでなく、データセット・ソフトウェア・著作権など、結果をもたらすツールが含まれるようになってきた。実験の動画や新規に構築したデータセットを投稿する形のジャーナルも生まれており、再現や再利用できる形のデータの構築および公開自体に学術的価値が認められつつある。例えば、Journal of Visualized Experiments は、実験の様子を動画として投稿するオンラインの論文誌である¹³⁾。また、Nature 社は Scientific Data という科学的に貴重で新規なデータセットを掲載する形のオンラインジャーナルを創刊し、データ再利用の時代が到来していることを強調している¹⁴⁾。従来どおりテキストや図表が主体の論文でも

オープンアクセス化（ウェブ上に誰もが無料で閲覧可能な状態に置くこと）が進み、例えば生物医学系の論文ではオープンアクセス化の割合が50%を超しているとのことである。この結果、論文レポジトリがSNSのように自動的にできあがるMendeleyのようなサービスができあがり、既存の文献データベースともリンクし始めた¹⁵⁾。また、研究成果の評価がアカデミアからのレビューのみであったことへの反省から、社会一般からの反応をリアルタイムで見ることができるAltmetricsという研究評価の仕組みも生まれている¹⁶⁾。

データの蓄積・公開がもたらす変化の1つに、研究活動のクラウドソーシング化の始まりがある。データ公開によるクラウド解析の例として、NASAは50年以上にわたって撮影した宇宙からの観測画像を公開し、分析作業に一般市民の協力も呼びかけている¹⁷⁾。多様な人々が参加すれば、これらが宇宙以外の研究にも役立つ可能性が高まる。データ収集のほうにもクラウドソーシングのトレンドがあり、例えば、水中生物観測では、クラウドファンディングで資金を集めた遠隔操作型水中探検ロボットの研究開発コミュニティが世界各地の海や湖で観測を行い、バーチャルなラボを形成するようになった¹⁸⁾。オープンデータポリシーについては本誌6月号でも述べたが、高等教育の知識共有にもフラット化が進展し、大学の講義が無料提供され始め(MOOCs)、世界中の高校生や中学生が有名大学の教養講座を受けられるようになった。論文のオープンアクセス化によって研究者でなくとも多くの論文を自由に読めるようになっており、立派な研究成果を挙げる10代の科学者が出てくるようになった。米国では大学の図書館や公立の博物館で「まるごとデジタル化」がかなり進行してきているが、将来的には出版・査読評価などの多くの機能を統合した科学サイト(Google Science?)が立ち上がるのではないかと憶測も出ており、研究環境に関するビジネスモデルは今後も変化していくと予想されている。

長期的には、知識社会が構成されるようになると、科学の知識は特定の国や組織のものではなく、アクセスの自由度という意味では、科学者と一般人の情報格差も、先進国と新興国

の情報格差も小さくなっていく。教育や研究という行為の場が教室や研究室のような閉ざされた空間からオープンな環境へと移行し、オープンエデュケーション・オープンサイエンス時代になっていく。教育の重点は、知識の伝承ではなく、これらの恩恵を受けられる能力の向上に移るはずである。

現在は、まだデータ自体に商品価値を認められる時代であるが、将来的にはそこから生み出される経済的価値や社会的価値のほうに注目が高まるはずである。我々はデータを利用すると同時に位置情報や嗜好などの情報を無意識のうちに提供しているが、性善説に立つならば、これを情報の民主化と考えることもできる。特に公的投資による研究開発においては、データ公開が研究開発投資に対する義務と位置づけられる方向で進んでいる。

6. ビッグデータの構造変容と価値創造³⁾

データは構造でみると「構造化データ」と「非構造化データ」に大別でき、両者とも量的存在感・リアルタイム性・多様性の3Vが増大していく。

「構造化データ」とは、データベースのように一定の形式で格納され、統計的に扱えるような形のデータを指す。従来の統計学は既存のデータベースを中心とする構造化データが分析対象であり、現時点では構造化データのうちサイズが大きいものをビッグデータと捉えている人も多い。現在のデータサイエンティストとは、主にこのようなデータ解析に力を発揮できる人を指している。データのタイプとしては、統計学で伝統的に扱ってきた数字データのほか、最近ではテキストデータからの意味解析(テキストマイニング)も注目され、市場調査などで生かされている。

構造化データは、相対的には正確性が高く、一定の水準を保ちつつ、システムティックに拡大していく。例えば、本誌9月号でも紹介されているように、病院では電子カルテやレセプトデータのような医療情報が日々蓄積され、医療に価値を与えようとしている(本誌9月号参照)。また、新たなデータ蓄積の例としては、例えばゲノム解析技術の進歩により単位時間あたりの遺伝子解析データ蓄積量は5年で1万

倍以上に拡大している（本誌7月号参照）。コストは10年間で10億分の1程度まで低減し、特定種のゲノム解析といった基礎的な解析研究は、個々人の遺伝子情報を提供できるサービスへと変化してきた。1人1人の遺伝子情報でも十分に大きいデータと言えるが、遺伝子検査サービスの普及により、すでに数十万人という単位で遺伝子情報が蓄積されている。同じフォーマットで蓄積されたデータからは、新薬や新治療法の可能性だけでなく、健康食品や健康維持対策といった医療関係とは異なる業界のビジネスチャンスが生まれている。欧米でこれから進められようとしているBrain Initiativeといった包括的な脳研究計画も、将来は同じようなビジネスチャンスを生むデータを提供していくと考えられている。

複数の構造化データが関連づけられると、さらに大きな価値が生まれやすくなる。例えば、マイナンバー制度によって医療・福祉データが他の社会保障関連データと結び付けられれば、社会的な意味で極めて大きな価値を見出すことができる（本誌9月号参照）。

また、構造化データからの知見がリアルタイムで公開されることによって、人の行動を変えることができる。簡単な例で言えば、雲の形で天候の変化を予知できることは、以前は農業に携わる人の特殊な能力だったかもしれないが、今では雨雲の観測データとして誰もがリアルタイムで知ることのできるデータとなっており、それによって多くの人々の行動パターンが変わっている。このようなリアルタイム情報提供は、渋滞の回避や非常時の安全管理などにおいて、大きな効果を発揮する。

一方、「非構造化データ」とは、文書・画像・センサー信号など、形式も格納場所も様々なデータである。コンピュータの利用範囲が広がるにつれて、テキスト・動画などが位置情報を伴ってカスケード的に急増し、また、各機器からのシグナルが広範囲にリアルタイムで集計され得るようになってきた。将来はこのような非構造化データがデータの大半を占める構造になっていく。特定の視点だけから見れば、非構造化データは正確性の保証されないデータを多く含み、価値密度の極めて低い、言わば「ごみデータ」の山であるが、そ

こには意識されていない潜在的な価値が存在する可能性がある。欧米のビッグデータのイニシアティブが将来的に期待しているのは、むしろ、このような非構造化データの増大とそのネットワーク化によって生まれる価値のほうである。将来のデータサイエンティストは、非構造化データのなかから新たな価値を見出すこと（これをキュレーションという）が主な仕事になっていくだろう。もっとも、非構造化データから新たな知を導くのは人工的な知能かもしれない。例えば、機械学習でも多くのデータから抽象的な概念を導ける可能性が出てきており、深層学習（Deep Learning）などと呼ばれ、注目されるようになっている。

非構造化データの規模が構造化データより膨大になっていく理由には、ウェブ上のデータ発信と収集がさらに容易になること、モバイル機器からのデータ収集や種々のデバイスからのデータ収集が自動化されること、そして大量データを扱える蓄積・処理技術が進展しつつあることなどが挙げられる。これまでも、ウェブ上の文書や画像などを周期的に取得し、自動的にデータベースに収集する回収プログラムにより、検索という概念が発達してきた。SNSの拡大に加え、大容量の映像データのサイトへの投稿がネット上で急増し、画像検索や音楽検索の技術が進展している。また、GPSや監視カメラの整備により、位置情報を含む画像データが利便性と社会の見守りのために自動蓄積されている。そこには、あらゆる「モノ」や「サービス」をネットワーク化するという考え方である、「モノのインターネット（Internet of Things：IoT）」や「サービスのインターネット（Internet of Services：IoS）」の具体化と進展がある。これらを目標に、生活環境に存在する多くの機器から信号データが常時集められる状況に移行しつつある。もちろん、ある種の非構造化データを一定の形式で格納すれば構造化データとなり、それらから新たな知見も得られるわけであるが、IoTやIoSのようなネットワーク化状態になると、我々が個々の構造化データにアクセスすることなしにシステムが自動的に最適状態に移行したり、非常時対応として自動的に意思決定がなされ、危機を最小限に

抑えることができるようになる。前述のドイツの Industrie4.0 とは製造業の現場をこの状態にしようというものであり、このような状況が関連企業やその他の施設にまで及ぶことまでを総称して、製造における Cyber-Physical Systems と表現している。また、スマートグリッドはエネルギーの IoT 化、すなわち Internet of Energy (IoE) を実現しようという概念である。

これまで計測されていなかったものから新たなデータ群が生まれ、想定外の相関が見つかる例も出てきている。例えば、人の状態や行動をウェアラブルな機器などで計測する技術 (Quantified Self) や脳とコンピュータの接続技術 (Brain-Computer Interfaces) は、メンタルを含む健康状態のチェックという意味だけでなく、行動経済学・組織経営学などに大きな知見を与えようとしており、最近の World Economic Forum (WEF) でも注目技術として採り上げられている¹⁹⁾。先導的な研究結果によれば、統計的には、人の行動パターンは比較的シンプルに表現できることが判明しつつある。例えば、暗黙知や技能として特定の人に帰属している特性をより客観的に形式知にすることができれば、組織の生産性向上や事業継続のための要素として極めて強力なものとなる。将来的には、ビジネスの生産性向上要素のなかに、個々の人の状態や行動に関するリアルタイムデータが組み入れられると予想される。人の行動や意識を、アンケートやインタビューなどの主観的・定性的・事後的データではなく、客観的・定量的・リアルタイムデータで表現できれば、本人も気づいていない失敗要素・事故リスク・過労などを事前に回避できる可能性がある。これまでの産業界では、「暗黙知」を人から人への伝達でしか伝えられないものと考えてきた。あくまで仮説ではあるが、後世から見れば、20 世紀の状態は印刷技術が現れる前に口述中心で技能伝達していた中世の状態に近かった、と見えるのかもしれない。

一方、非構造化データの処理を進展させるためには、巨大なデータセットが様々なコンピュータ上に分散して格納されている状況下でデータ処理を並列的に行うことが有効であり、世界規模の開発貢献者コミュニティによってボラン

ティアの技術開発プロジェクトが進められている。データの構造変化に追従する処理技術の開発は、これからこのようなオープン開発のスタイルが中心となって進展していくことだろう。その結果、形式の異なるデータが関係性をもちはじめ、予想外の相関が次々と発見され、次第に過去・現在の知見が将来予測へとつながっていくはずである。

しかし、データサイエンティストや情報ビジネス関係者以外のほとんどの人にとって、より関心を寄せるべきは、データ処理技術ではなく、ビッグデータから得られる価値のほうであろう。このような知を生かすには、その前提として、まずは判断する立場の者、すなわち組織を管理する者・戦略を立てる者・研究開発を進める者などに、データインテンシブ・データドリブンのマインドがあるかどうかのポイントとなる。残念ながら、データによって判断し、戦略を立てるという習慣は、少なくとも日本ではさほど根付いていない。また、多くの科学者・技術者は、観察・観測し、データを集め、その結果を議論して論文や報告を書いているわけだが、これからは、より大きなデータをより分かりやすく可視化し、そこから新たな知という価値を創出するといった科学の方法論も発達する。従来からの研究開発は、ややもすると観察や実験によって独自のデータを作成することに力点を置く傾向があり、学協会や研究会では個々の正確性や妥当性が主に議論される。もちろん、そのような観察・観測が必要な対象が減ったわけではないが、世界的な知の競争時代では、データの多くは「すでに存在する」あるいは「常に生まれていく」状態にあり、さらには極めて早く公開されるようになってきていることに、もっと目を向けるべきであろう。

また、蓄積・公開されたデータの利活用推進が可能かどうかは、おそらく多くの専門分野にとって情報スキルの獲得あるいは情報技術との融合的な取り組みができるかどうかにかかっている。日本では過度の規制や保護によるサービス業・農林水産業などでの ICT 技術導入の遅れ、医療データの未活用、政府の電子化の遅れなどが問題視されてきた。科学においても、バイオインフォマティクスなど

情報技術との融合的な研究がなかなか進まなかった過去の反省がある。アカデミアがビッグデータを活用できるかどうかは、これからの研究開発の成果に決定的な差をつけることになり得る懸念があり、特に人文社会科学系の研究の方法論に大きな影響が現れる可能性がある。世界の知の競争時代に乗り遅れないためには、専門領域にかかわらず、一刻も早く準備を始めなければならない。データから価値を見出せるキュレータ能力の醸成は、これまではあまり注目されていなかった教育要素であり、今後の人材育成や教育システムの重要な観点の1つになるであろう。

7. 「科学技術イノベーション政策のための科学」における対象データの変化

6月号に詳しく述べたように、「科学技術イノベーション政策のための科学」は、勘や経験に頼らず、より良い方針を決められるようにすることを意図して始められた研究開発プログラムである（本誌6月号参照）。したがって、データを重要視することは言うまでもないが、政策立案の根拠（エビデンス）となり得るものは何であるのか、あるいは、政策立案に対してより意味のあるデータは何であるかを見出すことが必要である。しかも、現実として、社会におけるデータ公開・利用の発展自体がすでに研究背景である社会の状況や各政策動向を大きく変えうるものになっている。

この研究開発プログラムの公募が始まった当初、政策立案の根拠（エビデンス）を提示するため、科学技術イノベーションに関する種々の計測が注目ポイントの1つとして挙げられた。その影響もあって、初年度である2011年度の選考では計量書誌学の要素を含むようなプロジェクトが多く採択された。特許や学術論文といった、さほどデータサイズは大きくないものの、比較的整備が進んでおり、市販のデータセットが入手できる指標の解析が多く提案で採り入れられていた。この頃すでに世界的にも計量書誌学は流行しており、誰でも研究対象にしやすいという面もあり、また、特許や学術論文数で世界における日本のシェア減少が国際競争力低下を表すもので

あると問題視されていたこともあって、科学技術イノベーションの度合いを示す指標として研究対象に採り入れるプロジェクトが数多く提案されたものと考えられる。図2の例を見ても分かるように、過去のトレンド検証や政策評価などには非常に有用であるが、世界的に見ると、これらのデータに対しては数量カウント中心の分析への注目は一段落し、フルテキストをマイニングして科学技術の潜在性や将来性を見ようという分析の発展が注目されている²⁰⁾。

最近の「科学技術イノベーション政策のための科学」の研究開発プロジェクト提案でも、従来の数量カウント中心の計量書誌学手法や母数に制限のあるアンケート調査などを用いる応募提案は減少し、社会的に意味をもつ大きなデータを扱おうとする提案が増えているように感じられる。このような傾向は、図2に示したビッグデータの研究の発展状況と対応しているようにも思われる。例えば、最近の応募提案では、資本や労働に関する経済統計、日本の状態を表す国政統計データ、人材に関わる調査データ、医療行為の結果を表すレセプトデータ、環境や資源の物流を現す種々の指標、エネルギー需給などの多様な構造化データが検討対象に入ってきている。このような新たな指標を用いて可視化を試みる研究提案が採択された例として、「市民生活・社会活動の安全確保政策のためのレジリエンス分析（プロジェクトリーダー：古田一雄氏、東京大学レジリエンス工学研究センター長）」という研究プロジェクト²¹⁾では、電力・ガス・水道・物流・通信など複数の重要インフラ相互の依存性を考慮に入れながら、脆弱性・耐性やリスクを可視化し、これらの重要インフラに関して回復能力であるレジリエンスの包括的評価手法と復旧プラン策定の判断支援手法を開発し、政府のレジリエンス向上策の立案や非常時対応のための組織制度設計を支援するための提言を行う計画である。また、これまでデータが離散している分野で、なんらかのデータセット、分析モデルの提供を試みるプロジェクトも採択されており、例えば、「リソースロジスティクスの可視化に立脚したイノベーション戦略策定支援（プロジェクトリ

ーダー：松八重一代氏、東北大学大学院工学研究科准教授)」という研究プロジェクト²¹⁾では、リンやニッケルといった特定元素を事例研究の対象として、サプライチェーンを通じた資源の流れを可視化し、新たな技術・システムの導入が社会・経済に与える波及効果分析を通じて、資源利用に関わるサプライチェーンリスクを回避するための意思決定支援マニュアルならびにデータベースの構築を行おうと計画している。

ただ、日本全体の状況として、政府のオープンデータ化には若干の進展が見られるものの、まだ、構造化データの拡大・公開・統合が政策立案に十分には生かされていない段階である。また、ビッグデータへの注目という意味でも、IT戦略などには現れているものの、イノベーション戦略のなかではあまり中心的な存在になっていない。したがって、上記のような研究開発プログラムでも、注目データの多様化が進んでいるとは言え、まずはこれまでの蓄積情報の可視化とその結果として導かれる政策提言が各研究開発目標となっている。また、非構造化データに関しては、残念ながら、まだこの研究開発プログラムのなかでは注目される段階に至っていない。将来に対応する準備としては、かなり心もとないように感じられる。

このような研究開発支援が行われにくい要因の1つとして、「研究開発の具体的目標を明確化しにくい」という特性があるように思われる。ビッグデータの解析というのは、どの程度の潜在性があるか分からない状態、つまり、「可視化されても、何が出てくるか分からない」という状態で、ひとまず始めなければならない。したがって、提案側では明確な具体目標を掲げて何かを解決するという形の提案は難しく、支援側でもその有望性の判断が難しい。米国のビッグデータ関連の研究開発も、全般的にハイリスク・ハイリターン型の研究開発として捉えられており、政府主導のイニシアティブの存在が無ければ、図2のような急速発展は起きなかったであろう。世界の注目が集まっているにもかかわらず、日本で取り上げられにくい背景には、ビッグデータの特性への理解が、産業界にも政府関係者

にもアカデミアにも十分に浸透していないという事情があるように思われる。

8. 終わりに

ビッグデータという言葉は一時的な流行語であるのかもしれないが、単に情報量が増大したという以上のことが着実に起こりつつあることを示している。すでに存在し、蓄積され、ウェブ上で公開されつつある膨大なデータの中から新たな価値の創出を図ろうとする世界の動きがあり、それがビッグデータという一見すると平凡な言葉が大きな関心を集める理由である。爆発的に増大するビッグデータを統計的に扱うことで、マクロにもミクロにも意味ある価値を見出そうとするデータ集約の技術は加速的に進展しているが、それらから得られた知に対して、次の行動が引き起こされつつあることにこそ、より大きな意味がある。それは生活スタイルや教育・研究スタイルを大きく変化させるものであり、社会全体がなんらかの影響を受けていく。

ビッグデータというキーワードの登場は、我々が知識社会への入り口に立っていることを示すものと考えたほうが良い。知識社会においては、膨大なデータのやりとりはインフラ化して、ほとんど意識されなくなるようになるのかもしれないが、見た目には何も変わっていないようでも種々の仕組みがまったく変化しているはずである。先進国の多くは人口動態の変化への対応が求められており、このような知識社会への変貌によって多くの適応策が図られていくはずである。

そこでは、ビジネスの内容も大きく変わっているだろう。もちろん、すべての現象がビッグデータで捉えられるわけではないだろうが、ごく一部の細部の議論を行なうよりは、俯瞰性のある捉え方で判断を行うほうが、全体の方向性を見間違える懸念は少ないだろう。現時点で個々の情報流通に着目するならば、情報セキュリティやプライバシーの保護など、情報自体に関する種々の問題は依然として存在し、法整備や規制緩和は追い付いていない。データ整備の点でもデータ取り扱いに関する政策の点でも検討課題はまだ多い。しかし、知識社会への移行は遅かれ早かれやってくるは

ずであり、そこではデータの利活用は個々情報のやりとりとは別次元のものと捉えられることになっているだろう。目前の問題にのみ目をとられていると、大きな変化の可能性を見過ごす懸念がある。少なくとも科学や教育においては、すでに世界は確実にオープンサイエンス・オープンエデュケーションの方向へ向かっている。この流れに沿って、次には世界の産業構造の大きな変化が誘発されることになるだろう。繰り返しになるが、ビッグデータは単に情報産業に価値をもたらすものではなく、第一次産業・第二次産業・第三次産業のいずれにも変化をもたらし得る。

このような観点から見ると、日本の産業界においても科学技術イノベーション政策においても、世界の動きに対応する準備が十分とは言えない現状が懸念される。

(おくわだ くみ)

注) 本稿の一部に、筆者による既発表の「ビッグデータからの新たな価値創造」³⁾の記述内容を加筆修正して用いています。

《参考文献》

- 1) 奥和田久美、過去の経験に頼ると危ない、という時代がやって来た、
<http://www.ristex.jp/stipolicy/topics/column/20121116.html>
- 2) P.F. ドラッカー、ネクスト・ソサエティ 歴史が見たことのない未来がはじまる、ダイヤモンド社、2002 年
- 3) 奥和田久美、ビッグデータからの新たな価値創造、生物多様性を規範とする革新的材料技術ニュースレター、Vol.2、No.1 (2013)
- 4) Smart Factory, <http://smartfactory.dfki.uni-kl.de/en>
- 5) ドイツ「Industrie 4.0」と EU における先端製造技術の取り組みに関する動向、日本貿易振興機構ブリュッセル事務所海外調査部欧州ロシア CIS 課、2014 年 6 月
- 6) 松田晋哉、日本の知的財産であるレセプト情報をいかに活用するのか、2014 年 8 月
http://www.kantei.go.jp/jp/singi/shakaihoshoukaikaku/chousakai_dai1/siryous5.pdf
- 7) ORCA サーベイランス、日本医師会総合政策研究機構、<http://infect.orca.med.or.jp/>
- 8) 症候群サーベイランス、国立感染症疫学センター、<http://www.syndromic-surveillance.net/>
- 9) 野村稔、米国政府のビッグデータへの取り組み、科学技術動向、2012 年 9・10 月号
- 10) 野村稔、奥和田久美、ビッグデータの研究開発推進の注目点、研究・技術計画学会第 27 回年次学術大会、2012 年 10 月
- 11) VALUENEX コンサルティング株式会社、Scopus に見るビッグデータの将来像 ―クラスター解析を活用した情報分析―、2014 年
http://www.elsevier.com/__data/assets/pdf_file/0008/198953/Scopus20140319.pdf
- 12) Microsoft Research, The Four Paradigm Data-Intensive Scientific Discovery, 2009, http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf
- 13) Journal of Visualized Experiments, <http://www.jove.com/>
- 14) Scientific Data, <http://www.natureasia.com/ja-jp/scientificdata/>
- 15) Mendeley Readership Statistics available in Scopus, 2014.3.7, <http://blog.scopus.com/>
- 16) Altmetrics Have Come of Age, Information Standards Quarterly, vol.25, No.2, 2013
- 17) data.NASA, <http://data.nasa.gov/>
- 18) OpenROV, <http://openrov.com/>
- 19) Top 10 Emerging Technologies 2014, World Economic Forum(WEF), 2014
http://www3.weforum.org/docs/GAC/2014/WEF_GAC_EmergingTechnologies_TopTen_Brochure_2014.pdf
- 20) Foresight and Understanding from Scientific Exposition (FUSE),
<http://www.iarpa.gov/index.php/research-programs/fuse>
- 21) 「科学技術イノベーション政策のための科学」プロジェクト紹介
<http://www.ristex.jp/stipolicy/project/>

* なお、本稿は、(独) 科学技術振興機構社会技術研究開発センター (JST RISTEX) による投稿論文シリーズの第 5 回目です。