

ビッグデータへの期待

Expectation for Big Data



一般社団法人日本知的財産協会 理事長

亀井 正博

1981年 富士通株式会社入社、2001年 同社 法務第一部長、2003年 同社 知的財産戦略室長、2010年 同社 知的財産権本部長、2014年 同社 法務・コンプライアンス・知的財産本部副本部長。2015年度 一般社団法人日本知的財産協会理事長

1. はじめに

ビッグデータという言葉が盛んに使われたのは、この5～6年のことだと思われる。ICT (Information & Communications Technology) 業界のマーケティングのための、いわゆるパスワードであると考えられる人も多いのではないかと思う一方で、ICT の発展によって、あらゆる情報がデジタル化され、あらゆる種類のデータが大容量に蓄積されつつあることを認識しない人もないものと想像される。

わが国では、ビッグデータという表現が使われたわけではないが、政府が関与した「情報爆発プロジェクト」(文科省 2005-2010)、「情報大航海プロジェクト」(経産省 2007-2009) など、ICT の利用によって発生する膨大なデータ(情報)を取り扱うための基礎技術、サービス等が研究された。最近では、あらゆる機器やセンサーがインターネットで接続されるIoT (Internet of Things) によって、膨大に収集されるデータを解析して得られた結果を現実世界にフィードバックするという「サイバー・フィジカル・システム」が実現しつつあることから、そのフィードバックは産業構造の変革をもたらすものとして、様々な取り組み、施策の検討がなされ

ている。データ分析の技術・手法としても、過去を分析するのみならず、未来予測のための分析、さらには最適解を得るための分析手法などが研究されている。産業界に身を置くものとして、さらなる技術の発展、またそれに伴ったビジネスの進捗が期待される場所である。そして、ビッグデータの分析によって得られる知見を活用することで、産業のみならず社会の「何か」を変えられるのではないかと期待感がある。

本稿では、ビッグデータの活用に関する筆者の個人的な期待と、併せて活用を促進するための法的・制度的環境に関する課題解決の方向性についての、やはり個人的な期待を述べてみたい。なお、ビッグデータの定義にはいくつかのバリエーションがあるようであるが、本稿では、一般的に言われる3Vの性質を有するデータ(の塊)というシンプルな定義のみを念頭に述べることにしたい。3Vとは、Volume: 大量、Variety: 多様性、Velocity: 高頻度・リアルタイム性である¹。

1 D. Laney "3D Data Management : Controlling Data Volume, Velocity, and Variety"、META Group (2001) <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

2. ビッグデータとしての特許情報への期待

知財関係者には、特許情報という馴染み深いビッグデータの例がある。大容量であり、また書誌事項のように定型化されてデータベースに収まっている情報のみならず、要約や特許請求の範囲、明細書、図面といった非構造化データを含むという点で、またデータをサンプリングするのではなくデータ全体を分析して活用されているという点から、ビッグデータの一つとして捉えてよいと思われる。そして、特許情報から「何か」を読み取ろうという取り組みはこれまでも多く行われてきた。様々

な分析ツールがベンダーから提供され、知財活動のパフォーマンスの測定、世の中の技術開発の動向把握、競合他社のベンチマーク、またライセンス、コラボレーション、企業買収の相手先を探すための一助とする等の目的で活用がなされている。例えば図1は、富士通(株)において、ある事業のコラボレーションの相手先を見つけるために役立つ事例²である。特許情報ビッグデータはこ

2 平塚信行『知財ビッグデータ活用による事業貢献～オープンイノベーション時代の知財活用』、FUJITSU Vol.65-5、p.9 (2014) <http://img.jp.fujitsu.com/downloads/jp/jmag/vol65-5/paper02.pdf>

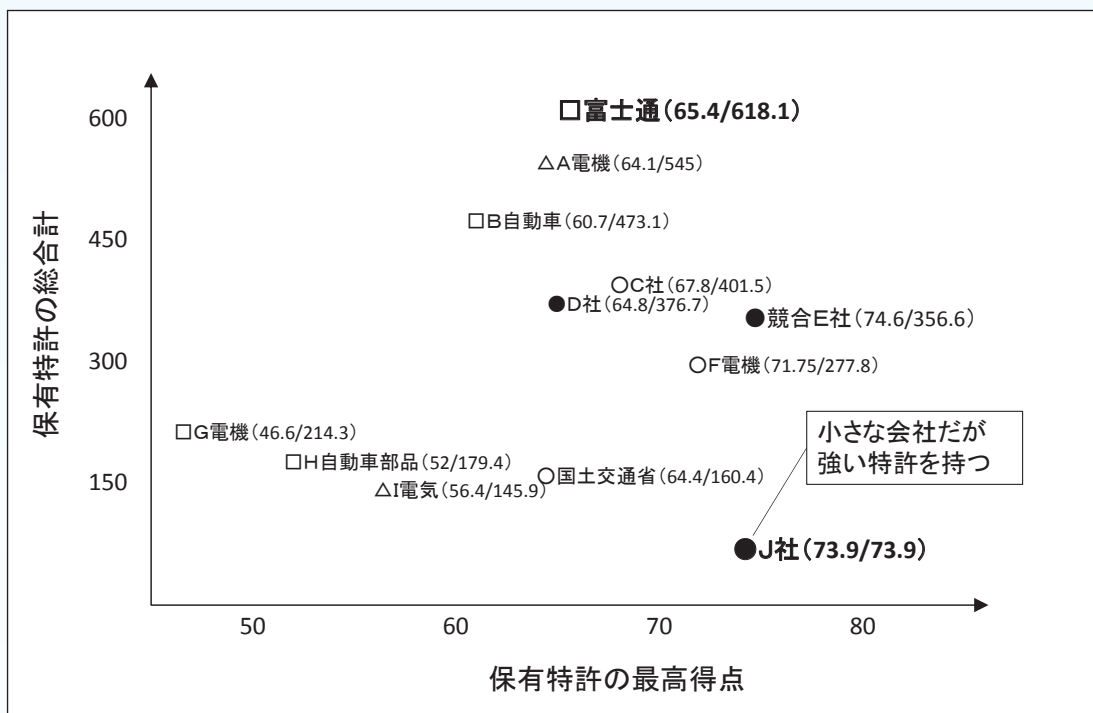


図1 注目企業・注目特許分析

れからも様々な見地からの分析が行われ、多くの知見をもたらしてくれるものと期待される。

ただ、分析対象として見たときに、特許情報は Velocity の点からは若干の課題が残っていると言える。特許制度は出願から 18 カ月後に公開されることから、公にアクセス可能な特許情報のデータベースでは、出願公開後の情報が収録される。しかし、その 18 カ月という期間のラグによって、特許情報の分析結果は「現在」ではなく、「少し前の過去」のものとならざるを得ない。技術進展の早い昨今では、世の中の変動に即して知見が得られるように、非公開状態の特許情報についても、ビッグデータ分析の対象とすることができるとよい。出願の内容そのものを公開するわけではないのだから、秘密を保持しつつ、例えば公益的な機関による一定の範囲での分析を可能にするなどの工夫はできないものだろうか。

3. 特許情報と他のビッグデータとの連携への期待

ビッグデータへの期待の本丸は、特許情報そのものの分析から得られる知見というだけではなく、特許情報と他の様々なデータとの組み合わせによって、いかなる知見が得られるのかという点にあるだろう。

ビッグデータの分析によって、いくつかの事象間の相関関係を知ることができるとされ、とりわけその因果関係に関わらず相関関係を導き出せるのがビッグデータ分析の特徴ともされている³。ビッグデータの利用によって、特許情報と、知財を取り巻いている環境に関する様々な情報との間で、これまでは思ってもみなかった相関関係を得られ、それを洞察することが可能となるものと期待される。

知財関係者が経営層からよく問われる質問の一つに、「これだけの特許資産を保有して、経営にどのような貢献をしているのか」というものがある。例えば医薬品のように製品と特許が非常に密接に関係しているような事業や、ライセンス収入による貢献など、一部の特許については定量的に答えることは可能だろうが、とりわけ価

値の顕在化していない特許を含む資特許産全体と様々な財務データとの相関関係の存在を指摘することはできていないのではないだろうか。産業セクターの違いによる事情があるかもしれないが、総じて言えば、特許資産の大きさ（例えば、特許の保有件数）と売上高や営業利益との相関関係があるのかというと、それはなさそうだというのが直観である。しかし、あるいはビッグデータの分析によって、何らかの相関関係を認めることができるのかもしれない⁴。また、特許資産の大きさと、その他のビッグデータと間に相関関係があるのかないのか。そうした分析については、まだあまり例が聞こえてこないが、例えば、研究者を取り巻く勤務環境や、研究者の日常行動と、知識創造の成果、例えば生まれてくる知財の質との間に何らかの相関関係を見つけることができるかもしれない。

そして、相関関係があるという結果を出発点として、その先には因果関係の探索が必要となる場合が多いだろう。とりわけ企業活動においては、どうしても因果関係を知って、最大の成果を得るための行動に落とし込みたいと考える。現象として把握される相関関係から、仮説によって因果関係を追求し、なすべきことを把握することになるだろう。上記の例では、知識創造環境を変えていく具体的なアクションにつなげ、その試行錯誤によって知財の質を高めていくということができる。

ここでは知財（特許情報）を中心に眺めてみたが、様々なビッグデータを組み合わせて分析することで「何か」を得るということでは、もちろん、知財中心の視点である必要はない。特許情報が、知財に関連しない「何か」をもたらしてくれる場合への期待もある。

4. ビッグデータ利用環境の整備への期待

4.1 公的機関・民間の保有データ活用

ビッグデータには、政府・公共機関が保有しているものがある。それらデータを開放して、民間のアクセスと二次利用を可能とすることは、産業振興のみならず

3 V. M = ショーンベルガー & K. キクエ『ビッグデータの正体 情報の産業革命が世界のすべてを変える』（斎藤栄一郎訳）、講談社（2013）

4 藤本祥二・石川温・水野貴之・渡辺努『ビッグデータを用いた特許の出願数と会社業績との関係』、金沢学院大学紀要「経営・経済・情報科学・自然科学編」第 12 号 p.57（2014）では、「（売上と出願数には）従業員数や資産の大きな大企業では大きくはないが相関はある」とする。

社会のイノベーションに有効である。2013年6月に閣議決定された「世界最先端IT国家創造宣言」において、政府・公共機関が保有するデータのオープンデータ化が掲げられた。その施策によって、この2年間で13,000を超えるデータセットが公開されたことは素晴らしいことである。今後も、より多くのデータの公開が進むことが望まれる。

他方で、民間の保有するビッグデータについても、ビジネスの観点からの積極的な公開という場合もあるだろうし、社会全体で相互活用してイノベーションに資するという視点からオープンデータ化されるものがあったとしてもよいと考えられる。民間の場合には、保有者の財産という側面も強く意識されるだろうから、次節に述べるような法的環境整備が進むかどうかにかかっているものと思われる。

勿論、共通する課題としては、データフォーマットやAPI (Application Program Interface)、セキュリティ技術等の標準化が進むことが肝要である。ISO・IEC等での国際標準化の進展にも期待される。

4.2 ビッグデータ利用のための制度的な整備

ビッグデータには、様々な性質のデータが含まれ得るが、利用を促進するには、それらデータの利用に関係する法制度の整備など、法的な側面からの環境を整えることも必要である。

最も分かりやすいのは個人情報保護に関してである。ビッグデータの活用促進とデータに含まれる個人情報の保護については適切なバランスが求められるところ、「個人情報の保護に関する法律」(個人情報保護法)が改正され、プライバシーに配慮しつつ利用できるように「匿名加工情報」が導入されることとなった。改正法の施行後に実際にビッグデータの活用が進むかどうかは、新たに設置される政府「個人情報保護委員会」での法運用や、データを活用しようとする産業界の法運用に依存しているものと考えられる。

他にも法的環境を整備すべきことがあろうが、本稿では、これから検討が進むだろう知財制度に関する検討課題について述べてみたい。

まずひとつ目の検討課題は、著作権法に関してである。公的機関や民間企業が保有するビッグデータもさること

ながら、インターネット上に膨大に書き込まれている情報も、貴重なビッグデータとしての価値を有している。そこには、映像、文章、音などの非構造化データが多く含まれるのであり、それらの多くは著作権で保護されるものだと想定される。著作物をデジタル化して取り込むことは、もはや非常に簡単な時代となっているし、非構造化データを分析するための技術も、今後、ますます進展していこう。その進展に応じて、非構造化データはビッグデータを構成する要素として活用価値が高まっていくものと考えられる。現行の著作権法47条の7では、著作権の権利制限として、著作物を「情報解析(多数の著作物その他の大量の情報から、当該情報を構成する言語、音、映像その他の要素に係る情報を抽出し、比較、分類その他の統計的な解析を行うこと)」のために複製や翻案をしてもよいと規定している。ビッグデータの分析は、多くがこの「情報解析」に該当するものと考えられるが、今後の社会的な要請や技術進展によって、該当するかどうか不明確となる場合も想定し得る。そうしたことを考えると、著作権の権利制限に少しの柔軟性を持たせる必要がある⁵。

この点に関しては、知財推進計画2015においても、「インターネット時代の新規ビジネスの創出、人工知能や3Dプリンティングの出現などの技術的・社会的変化やニーズを踏まえ、知財の権利保護と活用促進のバランスや国際的な動向を考慮しつつ、柔軟性の高い権利制限規定や円滑なライセンス体制など新しい時代に対応した制度等の在り方について検討する」としており、検討が進むことが期待される。

こうした権利制限規定の充実の一方で、ビッグデータを提供する際に、その利用の範囲を予め明確にしておくということは、実際にその利活用を進める効果がある。公的機関の保有するオープンデータについては、政府「データカタログサイト DATA.GO.JP」の利用規約では、データ提供者が著作権を有するデータについては、出典を明記すれば無償で自由に利用できるとされており、第三者に著作権があるものを除き、支障なく利用が

5 (一社)日本知的財産協会『世界最先端IT国家創造宣言に対する意見について』(2015) http://www.jipa.or.jp/jyohou_hasin/teigen_iken/15/150527_chosaku.pdf

できる。民間が保有するデータを開放する場合にも、利用態様は自由だが有償にする等の工夫と技術的な仕組み作りなどによって、活用が促進されていくことを期待したい。

ビッグデータが取引対象としての価値あるものとして捉えられるようになってくるのが、ビッグデータそのものの法的保護、とりわけ知的財産権による保護という考え方である。データベース全体を取引対象としたときには、全体を創作性あるものとして著作物であると法的に評価できるのであれば（そのような場合が多いと想定されるが）、多くの部分をコピーする行為には著作権侵害が主張できるだろう。また全体が著作物として評価できるものでなくとも、丸々全体をコピーされ営業行為をされたような場合には、不法行為を主張して損害賠償を求めることができるだろう⁶。問題は、ビッグデータを構成する要素としてのデータの保護である。個々のデータを保護したらどうなるかという頭の体操をしてみると、すぐに突き当たるのが、「データは誰のものなのか」という問題である。しかし、そのような議論はやめておきたいものである。個人情報について考えてみると、せっかく「匿名加工情報」によって活用が進むと期待されているのに、匿名化されてもなお、「この情報は自分のものである」という主張が可能となることになってしまう。個人情報でなくとも、著作物に該当しない数値や簡単な表現物等を知的財産権制度で保護することは避けておくべきだと考える。著作権法制度は、情報の保護と流通・利用を社会的にバランスするという根源的な価値判断によって、著作物でない情報は保護しないという制度である。取引価値があるからといって、そのバランスを変えてまで保護することが要請されているとは思えない。取引対象として権利を主張するには一方で義務を考えるべきでもある。誤ったデータ、瑕疵あるデータ、あるいは誤った予測によって、損害を被ったという人に対して、責任をとるのかどうか。上記の「DATA.GO.JP」では、無償・自由に利用できるかわりに、データの誤り等から生ずる責任について、公表者を一切免責している。

もちろん、すべてのデータを無償で供出すべきということではない。取引価値あるものとして尊重するととも

に、「データは誰のものなのか」議論の迷宮に陥ってしまっていて、その利活用が滞ることがあっては意味がない。

5. おわりに

本稿では、ビッグデータの活用を取り巻く、様々なレベルの期待を述べた。何のためにビッグデータを活用するのかという点からは、社会変革や産業振興のツールとして活用されることへの大きな期待がある。そしてビッグデータの活用を支える分析のために必要となる技術の進展、とりわけ非構造化データの分析技術の進展への期待感がある。活用に必要となる制度的課題への解決も期待したい。

最も肝心なのは、そうした技術や環境の下でビッグデータを利用して得られる知見なのであり、それは結局のところ、データから「何をしたいのか」という要求に始まり、システムが出力する相関関係等のアウトプットから、その「何か」とをつなぐストーリーやシナリオを語る利用者の問題である。分析して得られた相関関係を見ているだけでは先がなく、因果関係を追求し、また予測的な分析結果や指示的な分析結果を一助として、なすべき事を決めなければならない。つまりビッグデータを活用していくには、ビッグデータを分析できる人材、そして分析結果からなすべき事につなぐシナリオに落とし込むことのできる人材、総合的になすべき事を判断して決断できる人材が必要ということになる。知財関係でビッグデータを活用していこうとすれば、我々、知財関係者自身がそうした知識・スキルを身に着けるという課題設定もあろうが、少なくともそうした人材との協働を考えていかなければならない。様々なところでこうした、現状ではあまり多くはないと思われる人材の育成が進んでいくものと期待したい。

6 自動車整備業向けシステムのデータベースについての裁判例がある（東京地裁 2011 年 5 月 25 日判決）。

