

---

# 2014年総会シンポジウム「ビッグデータの可能性と課題 ——監視・シミュレーション・プライバシー」・論文

---

## ビッグデータと監視社会

### Big Data and Surveillance Society

キーワード：

ビッグデータ, 監視社会, プライバシー, 人工知能, 位置情報, 遺伝情報

keyword：

Big Data, surveillance society, privacy, artificial intelligence, location information, genetic information

神戸大学大学院人間発達環境学研究科 田 畑 暁 生

Kobe University, School of Human Development and Environment Akeo TABATA

---

#### 要 約

ビッグデータは情報社会における最新の流行語の一つとなり、日経や野村総研などが盛んに使ってそのビジネスを盛り立てているが、他方、ビッグデータ利用がもたらすプライバシー侵害問題についても、日立とJR東日本の事例のように注目を集めることがある。本論文では第1節でビッグデータの中身を再検討し、第2節でビッグデータによるプライバシー侵害問題の特徴を述べ、第3節では、いわゆる「監視社会」が、ビッグデータと人工知能技術の結びつきによって、人間の判断が機械に肩代わりされるような社会へと向かっていく可能性および危険性を論ずる。

#### Abstract

“Big Data” has become a “buzz-word” in information society, In Japan, this movement is driven by the NIKKEI and Nomura Research Institute. On the other hand, privacy problems are often discussed over Big Data, as of JR-East and Hitachi collaboration. This article reviews the meaning and content of Big Data in chapter 1, looks into privacy problem around Big Data in chapter 2, and thinks about surveillance society with Big Data in chapter 3. Surveillance society is now changing by “Big Data” and artificial intelligence (AI) technology.

## はじめに ビッグデータというバズワード

情報社会はそれ自体に関する流行語を次々と生み出してきた<sup>(1)</sup>。そうした言葉の中には、定着して日常語の仲間入りをしたものもあれば、ほどなく消えてしまうものもある。そして、「ビッグデータ」も、情報社会に関する最新の流行語、あるいは「バズワード」の一つといえるだろう。

「ビッグデータ」という言葉が記事で使われた頻度を主要4紙の新聞記事データベースで検索してみた結果が表-1である。2010年に始まり急速に伸びているのだが、特に日本経済新聞（およびその関連紙）での使用が突出していて、本紙のみに限っても、朝毎読3社の合計記事数を上回っている。今年に入ってから『日経ビッグデータ』の創刊からも分かるように、日経自体が「ビッグデータ」を煽る役割を主体的に行っている。

表-1 「ビッグデータ」を含む記事数

	日経系	うち日経本紙	朝日	毎日	読売
2010	1	0	0	0	0
2011	60	11	1	0	1
2012	376	108	9	2	5
2013	929	348	72	57	73
2014 (上半期)	555	213	63	54	41

西内(2013)は、流行になった「ビッグデータ狂想曲」を批判し、ビッグデータよりもサンプリング調査の方がコストが安く、また、データを解析する目標として、利益につながる「因果関係の道筋」の解明を挙げている。しかし翌年には西内自身が、『日経ビッグデータ』の広告に大きく登場しているのである。

野村総合研究所も、ビッグデータで金脈を掘り当てようとしている企業と言ってよいだろう。2011年に「ビッグデータ活用のロードマップ」を発表、日本で出版されたビッグデータ関連本の

多くに野村総研が関わっている<sup>(2)</sup>。

しかしビッグデータの定義は確立していない。大量のデータであることは必要条件であるだろうが、量的に明確な線引きはない。ダベンポート&キムは「並みはずれて大量なデータ、または、構造化されていないデータを意味する」とし、例として電子メール、フェイスブック、ヒトゲノムデータ、無線センサなどを例に挙げ（ダベンポート&キム、2014:p.17）、野村総研系の書物では「高精細、高頻度で生成され、多様性に富む（非構造的）データ」とし、また、Volume（量）、Velocity（速度）、Variety（多様性）という3つのV（さらにはそこにVeracity（真実性）を加えた4つ、Value（価値）を加えた5つのV）で始まる性質を特徴ととしている。金融機関や流通企業が従来から扱っているデータも、量としては相当大きなものであり、それと区別するため、「高頻度の更新」や「多様性」が言われる、ということもあるだろう（日立ホームページ内「ビッグデータへの道」第1回「ビッグデータとは」）。

平成26年度版『情報通信白書』では、ビッグデータの活用によって約60兆円の売上増があったとするのも、ビッグデータの定義があいまいなために、思い切り広く解釈して、ブームをあおったのだと思われる。個人情報保護法の改定も「ビッグデータ」を企業が（当事者の同意がなくても）利用したいという思惑に由来する部分が多い。

しかし定義があいまいとはいっても、日経や野村総研が盛んにおおっていることからわかるように、その使われる文脈がビジネス中心であるとは言えると思う。もちろん公共的な目的も存在する。一例が災害対策で、NHKが主導した「震災ビッグデータプロジェクト」は、POSデータ、交通データ、SNSでの書き込みなどを利用して震災後の人やモノの動きを鮮明に描き出した（阿部ほか、2014）。しかし、公共目的とは言っても、医療にせよ災害対策にせよ、効率や効用などのビジネスの言葉で論じられ、たとえば「ビッグブラザー」

という言葉が持っているような、権力関係についての文脈は通常、消去されている。

## 1 ビッグデータの中身

ビッグデータの中身はどのようなものだろうか。典型的な例をいくつか挙げると「1. 位置情報」、「2. SNS (ソーシャルメディア) などで発信される情報」、「3. 検索情報」、「4. アプリ利用情報」、「5. 医療情報・遺伝子情報」といったものが想定されている。言及は多くないが、「6. 監視カメラの映像」も、大量の構造化されていない映像情報という点でビッグデータの性質にあてはまっているだろう<sup>(3)</sup>。

特に1の位置情報は、まさにビッグデータの代表格といえるだろう。ヒトが動いた軌跡を、モバイル機器などが継続的に発信する。

ヒトの「位置」に関する情報は従来、「本籍」「住所」といったもので、「本籍」は戸籍簿のありかを示すパーチャルなインデックスに過ぎず、住民表における住所も、基本的にはそのヒトの寝所なり家産の置き場所であろうが、実質的に人がそこに住んでいるとは限らない<sup>(4)</sup>。

表から見える自動車のナンバープレートなども、その所有者の居住地域を示している。住所を知られたくないと、世田谷区の住民が、世田谷ナンバーの品川ナンバーからの自立に反対する、という事例も発生している。

2のソーシャルメディアをビッグデータとして利用する方法については、鳥海(2014)が、①「ソーシャルメディア理解としての利用」、②「社会を映すセンサーとしての利用」、③「データベースとしての利用」の3点に分けて整理している<sup>(5)</sup>。東日本大震災の災害対策としての利用は②にあたる。

3の「検索情報」、すなわち、グーグルなどの検索エンジンにどんな単語を入れて検索したかも、その人の欲望のありかをよく示すであろうし、

また、4のスマホ等のアプリについて、どのようなソフトをどんな風に利用したかも、その人となりを示すであろう。英単語アプリを使っている人には、英会話の広告は有効かもしれない。ゲームで課金されている人は「情弱」かもしれない。いわゆる名簿屋がもっとも欲しがるのも、騙されやすい人をあつめた「情弱」情報なのである。

5の医療情報はもともと、「機微情報」に属する 경우가多く、公共的性格が強い。政府は、医療費の抑制を目的に77億件以上のレセプトの「ビッグデータ」から、「レセプト・特定検診情報データベース白書(仮称)」を毎年編集・公表することを検討している(読売新聞2014年8月25日付け記事)。

人間の形質や行動の多くは、幸か不幸か、遺伝子によって直接規定されているわけではないが、性別や、血液型、ある一部の病気(たとえばハンチントン舞蹈病、テイ・ザックス病)などは、遺伝子によって強く規定される。さらに当人のみならず、血縁者についてもその近い度合いに応じて、病気のなりやすさなどの情報を与える。

遺伝子解析サービスは、米国では「23アンドミー」社が中心であり(ライバルであったデコード・ジェノミクス社やナビジェネティクス社は、2012年にそれぞれ買収され、個人向けのゲノム解析サービスを停止した)、日本でも各種の医薬系の会社で始まったが、2014年にはヤフーやDeNAといったIT系企業が参入し、話題を呼んだ。

## 2 ビッグデータとプライバシー

プライバシーに特に敏感な人々は、ビッグデータの利活用に素早く反応する。その代表的な事例が、JR東日本と日立が組んだ、乗降客データの分析・販売にまつわるものであろう。2013年、JR東日本がSuicaから得られた情報を、日立が解析して「駅利用情報分析レポート」として提供し

よとのサービスが厳しい批判を受け、対応として設置された堀部政男氏らをメンバーとする有識者会議（2013年9月から12月、2014年2月に中間とりまとめを発表）で、データの社外への提供は、個人識別性がなく、たとえ個人情報保護法等に違反していなくても、一層慎重に扱う必要があるとされ、JR東日本は、データの提供を見合わせる事となった。

JR東と日立の側では、匿名化したデータの分析・販売は個人情報保護法の観点から問題がないと考えたのだが、個人情報保護法の解釈にも曖昧なところがあり、匿名化したとしても個人識別される（非特定識別情報）のなら、個人情報に入る（容易に結合して個人を特定可能）との解釈も十分成り立つ。個人を特定できるのが「特定情報」、特定はできないが、ある一人の情報であると識別可能なものが「識別情報」とされるのだが、また、特定は多くの場合個人名をキーとするであろうが、個人名で一人に絞られるかどうかは、その名前が多い名前かどうかにも依存する。ありふれた苗字とありふれた名前では、同じ名の人が1万人を超える場合さえあり、逆に珍しい苗字の場合には名前だけで一人に絞られることがある。住基番号（や、旅券番号や免許証番号）は個人を特定するが、数字だけで特定されたというのは、人間の感覚では納得できないところもあるだろう。

「k-匿名性」「l-多様性」「t-近接性」といった、匿名化したデータからの個人特定を不可逆にする技術がさまざまに開発されてはいる。しかし、他のデータと結合することで、個人特定にいたる可能性は、ゼロにはならないだろう。AOLやネットフリックスが公開した、匿名化した検索データからも、個人の特定制がなされてしまったのである（ショーンベルガー＝クキエ、2013:231-233）。

総務省「位置情報プライバシーレポート」（2014）によると、「災害（緊急）被災地域の被害状況把握・救援活動・避難・帰宅支援」のためであれば、位置情報の利用を「許容できる」とし

た人が67.3%、「条件つきで許容できる」とした人が28.3%で、合わせて95%を超える（どんな場合でも許容できないとした人は4.6%）。また、「防災計画策定・避難場所設置・防犯」のためであれば、「許容できる」が58.3%、「条件付きで許容できる」とした人が35.5%で、これも合わせると9割を超える。それに対して、「広告・マーケティングやサービス向上」という理由の場合、「許容できる」は9.1%にとどまり、「条件付きで許容できる」の46.1%と合わせて、やっと過半数といった割合である。緊急事態であれば許容するデータ利用であっても、広告などのための利用は、多くの人が好ましくないと考えている。

心理学には有名な「ジョハリの窓」（ジョセフ・ルフトとハリー・インガムが考案）というモデルがある。自分も他人も知っている自己を「開放の窓」、自分は知っているが他人は知らない自己を「秘密の窓」と呼ぶ。プライバシー情報として意識されるのは、多くはこの「秘密の窓」に入る事柄だろう。しかしそれ以外にも、自分は気付いていないが、他人には知られている自己として「盲点の窓」、そして自分も他人も知らない自己として「未知の窓」という4つの部分を4象限に置いている。特にビッグデータには、意識しない自分の行動やゲノムといった「未知の窓」「盲点の窓」に属するような情報も、多く含まれているのだ。そうした分野の情報が「利活用」されることは、反対の声を上げる人は多くはないにしても、自分さえ意識しないようなデータが含まれるビッグデータの利活用には、多くの人が本能的（？）に、「気持ち悪さ」を感じているのかもしれない。

プライバシー侵害への懸念を、社会全体の問題に敷衍すれば、いわゆる「監視社会」化の問題とつながる。

### 3 監視社会の変質

「監視社会」もまた、情報社会のパスワードの

一つかもしれない。ビッグブラザー（オーウェルが小説「1984年」で描いた独裁者）、パノプティコン（ベンサム、フーコー）といった概念とともに、監視カメラ台数の増加、ICカードの普及、企業および国家による活発なデータ収集と利用、さらにはカメラ付きケータイの普及や一般人による活発なSNSへの映像投稿などについて、「監視社会化」として懸念が表明されてきた。ビッグデータの利活用の増大はまさに、量としての「監視の拡大」といえるだろう。

では、監視社会に質的な変容をもたらすものだろうか？

ビッグデータが監視社会に質的な変容をもたらすとすれば、それはデータの監視や解釈が、ますます人間の手を離れて機械で完結したものになっていく、ということだろう。例えば監視カメラの台数はますます増えているが、それを「監視する」人間の数は足りない。人間の頭数が足りているとしても、その仕事をさせるには、費用が発生する。人件費を減らして効率を上げるとの動きはますます加速するであろうから、上手な「ゲーミフィケーション」<sup>(6)</sup>を行えば、ひょっとすると失業者などの有効活用につながるのかもしれないが、その場合でも失業問題の解決にはならない。

ビッグデータの監視者や解釈者が人間から機械へと大幅に肩代わりされ、人間がそれにためらわずに依存するようになったとき、人間がそのままでは理解できないデータの「意味」は、ブラックボックス化した「御託宣」として受け取られるおそれがある。おうむ返しを基本とする対話型「精神分析プログラム」を作ったワイゼンバウムが、それにはまり込む人が多いのに驚いて研究を辞めたという有名なエピソード（ワイゼンバウム、1976=1979）を思い起こさせる。

小林（2013）が指摘するように、一時は研究が停滞する冬の時期があった人工知能（AI）研究に春をもたらしたのはビッグデータと言える。1980年代末、UCLAの科学者ジュディア・パー

ルが、「ベイジアン・ネットワーク」に代表される統計・確率的な考え方を人工知能研究に持ち込み、そうした考えに影響を受けた科学者・エンジニアたちが例えばグーグルのようなIT企業に入社して、例えば機械翻訳においても、それまでの論理とは違い、ネット上から大量のデータを収集することで、ありそうな言葉の組み合わせを探するという方法で、次第に精度を上げていった（小林、2013:75-77）。

『日経コンピュータ』誌2014年10月2日号、特集「ビッグデータは人工知能に任せた」では、人間は不要になるわけではないとし、人間の役割として「適切な手法の選択」「適用領域の発見」「機械による仮説の検証」「データの収集」の4点を挙げているが、このいずれも、人間が特に機械に対して優位性を持つ領域ではないだろう。

経済学者タイラー・コーエンの『大逆転』では「すでに科学分野によっては、人間の頭脳では対処できないくらい、理論や仮説が複雑化している可能性がある」（p.254）。「コンピュータは次第に、実質的な研究活動で中心的位置を占めるようになり、研究計画の設計でも主たる役割を担うようになる」（p.261）。「（フェイスブック、グーグル、アマゾンといった）ウェブ関連企業は、理論にはあまり関心を示さず、オンラインコミュニケーションを通じて得られたデータを統計学的に分析する。こうして生まれた「ビッグデータ」は、次のビジネス革命の担い手と期待されている」（p.264）と指摘する。

日立のエンジニアによる矢野（2014）には、「ビッグデータで儲けるための3原則」が出てくる。「第1の原則 向上すべき業績（アウトカム）を明確にする」「第2の原則 関係するデータを広く収集する」「第3の原則 仮説に頼らず、コンピュータに業績向上策をデータから逆推定させる」。矢野の挙げている事例は以下のようなものだ。

ある店舗の売り上げを向上させるために、二人の専門家と、日立の人工知能ソフトウェア（略称

「H」)とを競わせた。専門家はさまざまなヒアリングやデータから、店内広告の設置や棚配置の改善などを行った。「H」は、大量のデータを6000個の業績要因に分解して、純粋にデータとの相関関係だけから、店内のある特定の場所（高感度スポット）に従業員が立っていることで、購買金額が上がる（10秒長く立っていると、客単価が145円上がる）ことを発見した。そして、実際に店員がそこに立つ時間を増やすことで、売上げが上がった（専門家による助言は、ほとんど成果を上げなかった）。ここから矢野は、「人による仮説検証型分析はビッグデータに通用しない」とし、「仮説に頼らず、コンピュータに業績向上策をデータから逆推定させる」ことを提案しているのである。

また、前述したゲノム情報は単体では、基本的に変化しないものであるし、ビッグデータといえるかどうか微妙であるが、ゲノム情報と、その人がどんな食事をし、どの程度運動しているかといった位置情報や行動情報、さらにtwitterやFacebookなどでの書き込み情報を組み合わせれば、その人が特定の病気をいつ発症するか、かなりの程度で予測できるのかもしれない。それを分析し、余命の推計を高い確率で予測できたり、医薬や医療保険の販売などにつなげられるのかもしれない。

それどころか、さほど正確な予測などできなくても、血液型性格診断やら六占星術が世にはびこる（こうした事実は、いかに普通の人々が思考を節約して生きているのか、もしくは、思考のコストがいかに高いかを示しているように思われる）ように、人間をいくつかのパターンに分けて単純化するような、変な予言が独り歩きする危険がある。田畑（2003b）でも論じたが、遺伝子それ自体にも、文化的な物神性がまわりついているのである。こうした動きをとどめることは難しいであろうが、ビッグデータによる予測が行われる基本的な統計学的手法（たとえば重回帰分析）くらいは、万人が理解している必要があるのではなか

ろうか。

## おわりに ビッグデータを越えて

田畑（2003a）で指摘した、「監視の主体は誰か」「誰が解釈するのか」「監視の分散をどう評価するのか」といった論点の重要性は、ビッグデータという言葉がまき散らされようが、基本的には変化をしないだろう。

すでに述べたように、ビッグデータという言葉ばかりに目を奪われると、「ビッグブラザー」的な権力から目がそらされる危険がある。たとえば政治家や高級官僚、警察・検察官僚のような強大な権力者に関しては、汚職などの抑止のために、24時間のライフログを記録し、一定の時間後に公開するというのは技術的には可能だが、そうした話にはなかなかならない。それどころか、いつまでも秘密を延長できるような、政府に甘い「秘密保護法」が、批判を受けながらも可決成立してしまう。

政府や大企業が、高度な情報技術や機械的知能を使って一方的に人々を監視するとしたら、どれほど利便性を増してくれるとしても、あやうい社会であろう。判断のブラックボックス化がいやおうなしに進む中で、その判断メカニズムの在り方のみならず、ビッグブラザーやビッグシスターについての情報こそ、万人に対して開かれてあるべきなのではないか（もっとも、政治のトップが自前の判断能力を持たず、官僚等の実質的言いなりという可能性もあるが）。監視社会化が必然であるにしても、誰が誰を、どの程度監視するのかは多様であり得る。民衆が国家に対する監視を強化する社会を、監視社会と呼ぶこともできるのだ。われわれは、「より望ましい監視社会」を目指さなくてはならないだろう。

## 注

- (1) ニューメディア, マルチメディア, インターネット, ソーシャルメディア, ウェブ2.0, クラウド, ユビキタスなど, 枚挙にいとまない。こうした言葉は, 消えていったものもあれば, 定着したものもある。
- (2) 鈴木 (2011), 城田 (2012), 安岡編 (2012), 野村総合研究所 (2012), 小林 (2014) など, いずれも, 著者や編者は野村総研の社員なのである。
- (3) 監視カメラ映像がビッグデータとして意識されることが多くないのはおそらく, 事前に個体識別されていない(顔認証や各種「生体認証」で識別可能), 権利関係が(特に)不明確, 各種カメラの所有者がまちまち, などのためであろう。現在でも, 事件後に監視カメラによる映像は捜査機関によって分析され, 犯人検挙に役立てられてはいるが, 監視カメラの精度がさらに上がり, 顔認証技術がさらに発展を遂げれば, 監視カメラの映像だけから「それが誰か」を機械自体が特定する可能性は, それが社会的に容認されるかどうかは別として, 大いに高まるだろう。
- (4) 住民票の住所を変えずに実質的な転居を行う人はいる。例えばDVで家族から逃げたいという人にとっては, 追われる危険から実態に合わせた住所変更ができず, 「無戸籍児」の問題を生んだりした。尼崎連続殺人事件の犯人グループが, 逃げた「家族」の動向を住民票で常にチェックしていたことは, 戦慄を与えた。亡くなった高齢者の年金を家族が詐取するために, 死亡届を出さないでいる事件も, 各地で発生した。
- (5) 田村 (2014) は, SNSなどの情報を企業が利用することは, 消費者にとって利益になるとは限らないことを指摘する。「フェイスブックに代表されるSNSでは個人の属

性や興味関心, 交友関係などの情報があふれており, Suicaなどの電子マネーの利用履歴からは個人の行動パターンや購買行動の情報が収集される。そういった情報を企業が市場の細分化のために利用することの効果は必ずしも自明ではなく, 消費者余剰の改善のためにどのような利用が許されるべきか個別具体的な議論が必要である」(p.96)。

- (6) ゲーミフィケーションとは, ゲーム化することで, 人間を一定の行動へと楽しんで向かわせることを指す, 近年注目されている技術である。ゲーミフィケーションを使って, スピード違反を大幅に減らした成功例が, 野村総合研究所 (2012) に掲載されている。ストックホルムで行われた「Speed Camera Lottery」では, 監視カメラの前で, スピード違反したドライバーの罰金を, 法定速度を守ったドライバーに抽選で分配する, というもので, 3日間の実証実験中, 速度は平均して22%低下した(pp.062-063)。

## 文献

- 安部博史+NHKスペシャル (2014) 「震災ビッグデータ」制作班編『震災ビッグデータ』NHK出版。
- Tyler Cowen (2013) “Average is Over” Dutton Adult.=タイラー・コーエン, 池村千秋訳 (2014) 『大格差』NTT出版。
- Craig, T and M.E.Ludoloff (2011) “Privacy and Big Data”, O’Reilly.
- Davenport, T.J. & K. Jinho (2013) “Keeping Up with the Quants” Harvard Business Review Press.=トーマス・ダベンポート, キム・ジノ, 古川奈々子訳 (2014) 『真実を見抜く分析力』日経BP。
- Halavais, A (2009) “Search Engine Society” Polity=アレクサンダー・ハラヴェ, 田畑暁生訳

- (2009)『ネット検索革命』青土社,
- 小林慎太郎 (2014)『パーソナルデータの教科書』日経BP.
- 小林雅一(2013)『クラウドからAIへ』朝日新書.
- 西内啓 (2013)『統計学が最強の学問である』ダイヤモンド社.
- 野村総合研究所 (2012)『ビッグデータ革命』アスキー・メディアワークス.
- 坂井豊貴 (2014)『メカニズムデザインと意思決定のフロンティア』慶応義塾大学出版会.
- 嶋田茂ほか (2014)『ビッグデータ・マネジメント』エヌティーエス.
- 城田真琴 (2012)『ビッグデータの衝撃』東洋経済新報社.
- Schonberger,V.M. &K,Cukier (2013) “BIG DATA”Houghton Mifflin Harcourt.=ショーンベルガー他, 斎藤栄一郎訳 (2013)『ビッグデータの正体』講談社.
- 鈴木良介 (2011)『ビッグデータビジネスの時代』翔泳社.
- 田畑暁生 (2003a)「監視社会論の射程」『人間科学研究』第10巻第1号.
- 田畑暁生 (2003b)「社会情報としての遺伝子」『社会情報学研究』第7号.
- 田村彌(2014)「最適シグナル」坂井豊貴編[2014]所収, p.96
- 鳥海不二夫 (2014)「ソーシャルメディアにおけるビッグデータ処理」嶋田ほか (2014) 所収, Weizenbaum,J.,”Computer (1976) Power and Human Reason”W.H.Freeman Company=ジョセフ・ワイゼンバウム,秋葉忠利訳 (1979)『コンピュータ・パワー』サイマル出版会,
- 矢野和男 (2014)『データの見えざる手』草思社.
- 安岡寛道編 (2012)『ビッグデータ時代のライフログ』東洋経済新報社.
- 『日経コンピュータ』2014年10月2日号, 特集「ビッグデータは人工知能に任せた!」日経BP社.