

# 米国におけるビッグデータ活用に関する動向

八山 幸司  
JETRO/IPA New York

## 1 はじめに

高度に情報化した現代社会では、技術的革新により社会や経済も新しい形へと変化しており、人々のコミュニケーションやビジネスの場の多くがインターネットへと移行してきている。そのような状況の中、人々の生活から日々生み出されるデータは社会の動きそのものを表しており、データを通して社会をマクロ的な視点で分析し、その中にあるトレンドや新しい価値を見出すビッグデータが社会に新しい変革をもたらそうとしている。例えば、多くの企業では競争力を保つためにビッグデータを使用した戦略が必須となっており、従来のビジネス戦略は、データ重視型 (data-driven) やアナリティクス重視型 (analytics-driven) へと変化してきている<sup>1</sup>。ビッグデータを活用する技術も、現在のトレンドや過去の分析だけでなく、未来の価値を見出す予測分析 (predictive analysis) などへと発展している<sup>2</sup>。ビッグデータを貴重な資源と捉え活用が広まる一方で、プライバシー問題への懸念などが出てきており、政府、企業、民間のあらゆるレベルで取り組みが必要とされている。今号では、米国におけるビッグデータの活用とプライバシー問題への取り組みについて紹介する。

最初に、急成長するビッグデータ市場について紹介する。2014 年のビッグデータ市場は前年から引き続き急成長が続くと見られ、またビッグデータ関連製品・サービスの市場は様々な方面で成長していくと見られる。それらに伴い、企業によるデータ分析に特化したデータサイエンティストの需要も高まっている。ビッグデータを扱ったベンチャー企業への投資も大きなものとなってきており、10 億ドル規模の投資を受けている企業や、医療アナリティクス分野で設立直後から投資を受けるベンチャー企業も登場している。

次に、連邦政府によるビッグデータに関連した政策を紹介する。2014 年 5 月に連邦政府から発表されたビッグデータに関する報告書では、政府によるビッグデータに関する取り組みを評価し、様々な観点から今後必要な取り組みが述べられている。ビッグデータ研究開発イニシアチブでは、2012 年に連邦政府から 2 億ドルの投資が行われているが、2013 年には産学官のビッグデータパートナーシップによる共同研究プログラムが始まっている。

更に次の章では、ビッグデータを新しく取り入れる分野に焦点を当て、医療、教育、社会サービスにおけるビッグデータの活用を紹介する。医療分野でのビッグデータの活用ではウェアラブル端末とビッグデータの組み合わせによるパーキンソン病への取り組みや、人工知能を使ったデータ解析による心臓病の早期発見の試みがある。教育分野では、ビッグデータを利用した学習プログラムなど、様々な活用事例を紹介する。また、教育分野におけるビッグデータ活用のために多大な支援を受けて立ち上がった inBloom<sup>3</sup>は、生徒のプライバシーを心配する保護者からの非難などによりサービスを継続できなかったケースとなっている。社会サービスでは、子供の虐待やホームレス問題への解決にビッグデータの分析が使われている。

最後に、ビッグデータを活用する上で最も懸念される、プライバシー問題への取り組みについて紹介する。データ収集と分析を行うデータブローカーは、ビッグデータを使った経済活動で重要な役割を果たしている

<sup>1</sup> <http://www.forbes.com/sites/louiscolombus/2014/06/24/roundup-of-analytics-big-data-business-intelligence-forecasts-and-market-estimates-2014/>

<sup>2</sup> <http://www.informationweek.com/big-data/big-data-analytics/big-data-analytics-descriptive-vs-predictive-vs-prescriptive/d/d-id/1113279>

<sup>3</sup> ジョージア州アトランタを拠点とする非営利団体。高セキュリティの教育データ保管サービスを教育機関に提供していた。  
<http://www.nytimes.com/2013/10/06/business/deciding-who-sees-students-data.html>

が、同時に政府から、その活動に対して法規制を必要とする声も出ている。また、プライバシー問題に対し州政府による法規制も進んでおり、カリフォルニア州やフロリダ州ではインターネット時代に即したプライバシー法が制定されている。EU で認められた「忘れられる権利」では、Google 社などが対応を迫られており、プライバシーと言論の自由の間でどのようなバランスが必要か議論は続いている。

米国におけるビッグデータは、政府による投資や企業の積極的な活用によって急速に発達しているといえるが、プライバシー問題など対応を迫られている面もある。米国の取り組みを見ることで、ビッグデータによる恩恵だけではなく、日本で対応が必要な問題についても考えるきっかけにしていきたい。

## 2 ビッグデータ市場

### (1) 拡大するビッグデータ市場

ビッグデータの定義には様々な解釈があるものの、コンサルティング企業の McKinsey & Company 社によると、「従来のソフトウェアでは取得、保存、管理、分析が出来ないほどのサイズを持ったデータ」と定義されている。また、この定義の基準が主観的なものとなっているのは、巨大なデータと判断するための基準が変えられるように意図したものである、と付け加えられている<sup>4</sup>。

ビッグデータの活用が飛躍的に成長すると見込まれる中、市場規模も大きく成長すると予想されている。2014 年の世界のビッグデータ市場は 285 億ドルに達すると見込まれており、2017 年には 501 億ドルにまで成長すると予想されている<sup>5</sup>。2014 年度ビッグデータ市場の成長率は前年よりわずかに下落するものの 53%の成長率になると見られており、2017 年までの年間平均成長率は 38%になると見られている。2013 年のビッグデータ市場では、サービス関連が全体の利益の 40%を占めており、ハードウェア関連が 38%、ソフトウェア関連が 22%と続いている。これはビッグデータがソフトウェアや専門家によるサービスの提供を中心としたビジネスモデルとなっているためである。ビッグデータ市場の急成長を牽引している要因として以下のようなものが挙げられる<sup>6</sup>。

- 大手 IT 企業やビッグデータのサービス事業者が、政府機関や企業顧客などに対して、将来的なロードマップや市場予測に関する製品やサービスを積極的に提供している。
- ビッグデータに関連した製品やサービスがより成熟したものになってきている。例えば、大規模データの分散処理フレームワーク Hadoop や、クラウドベースのビッグデータサービスの進化などが挙げられる。
- ビッグデータの関連技術が企業や政府機関にとって、プライバシー、セキュリティ、データのバックアップや復元などの管理など、守りの技術としても重要視されている。
- ビッグデータ専門の企業が様々な企業と協力することが多くなってきており、ビッグデータを活用したサービスが身近なものになってきている。

より多くのビッグデータの活用やクラウドベースで提供するサービスが、今後のビッグデータ市場の拡大で重要な役割を果たすと考えられている。サーバーやストレージなどのインフラの整備が進んできていることもあり、今後はさらにビッグデータ関連サービスの普及が進むと見られる<sup>7</sup>。図表 1 はビッグデータの世界市場予測を表したものとなっている。

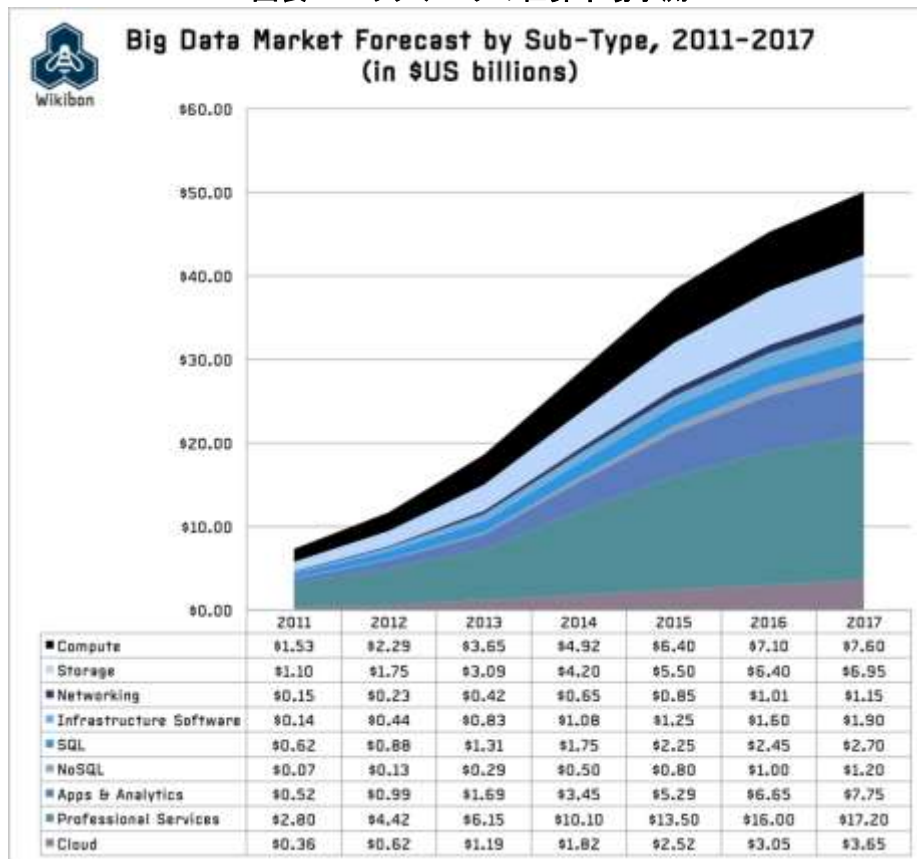
<sup>4</sup> <http://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/>

<sup>5</sup> <http://www.forbes.com/sites/louiscolombus/2014/06/24/roundup-of-analytics-big-data-business-intelligence-forecasts-and-market-estimates-2014/>

<sup>6</sup> [http://wikibon.org/wiki/v/Big\\_Data\\_Vendor\\_Revenue\\_and\\_Market\\_Forecast\\_2013-2017](http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2013-2017)

<sup>7</sup> [http://wikibon.org/wiki/v/Big\\_Data\\_Vendor\\_Revenue\\_and\\_Market\\_Forecast\\_2013-2017](http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2013-2017)

図表 1:ビッグデータの世界市場予測



出典: Wikibon<sup>8</sup>

ビッグデータ関連製品・サービスの市場は様々な方面に広がっており、以下のような市場やサービスの成長が期待されている。

- クラウドベースのビジネスインテリジェンスの成長が著しく、世界の市場規模は 2013 年の 7.5 億ドルから 2018 年には 29.4 億ドルに伸びると見られており、年 31%の成長率が見込まれている。また、ビジネスインテリジェンスのプラットフォームを提供する企業として SAP 社、IBM 社、SAS 社、Microsoft 社、Oracle 社、Information Builders 社、MicroStrategy 社、Actuate 社などが市場を牽引していくと見られている<sup>9</sup>。
- 予測分析(Advanced and Predictive Analytics:APA)ソフトウェアの世界市場は 2013 年に 22 億ドルの規模となっているが、2018 年までに 34 億ドルにまで成長すると見られており、年間 9.9%の成長が続くと見られている。現在の APA ソフトウェアは統計用ソフトウェアなどよりも使い勝手の悪い面があるものの、よりグラフィカルで使いやすいものになることで、ビジネス分析の市場に貢献すると見られている。2013 年の時点で市場におけるトップ 3 の企業は、SAS 社(市場シェア 35.4%)、IBM(市場シェア 17%)、Microsoft 社(市場シェア 3%)となっている<sup>10</sup>。

<sup>8</sup> [http://wikibon.org/wiki/v/Big\\_Data\\_Vendor\\_Revenue\\_and\\_Market\\_Forecast\\_2013-2017](http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2013-2017)

<sup>9</sup> <http://www.forbes.com/sites/louiscolumnbus/2014/06/24/roundup-of-analytics-big-data-business-intelligence-forecasts-and-market-estimates-2014/>

<sup>10</sup> <http://www.forbes.com/sites/louiscolumnbus/2014/06/24/roundup-of-analytics-big-data-business-intelligence-forecasts-and-market-estimates-2014/>

- 高性能データ分析 (High Performance Data Analysis: HPDA) に使われるサーバーの世界市場は年間 23.5% の成長率となっており、2018 年までに 27 億ドルに拡大すると見られている<sup>11</sup>。
- 大規模データの分散処理フレームワーク Hadoop を使ったソフトウェアの世界市場では、Amazon Web Services 社、Cloudera 社、Hortonworks 社、IBM 社、MapR Technologies 社、Pivotal Software 社、Teradata 社などが市場を牽引していくと見られる<sup>12</sup>。

## (2) 需要が高まるデータサイエンティスト

ビッグデータ市場が拡大する一方で、データの加工や分析を行うデータサイエンティストの需要が高まっている。数年前まではあまり注目されていなかったデータサイエンティストであるが、ハイテク産業を中心に、小売業、銀行、重工業、人材派遣業などで需要が高まっており、インターネットやスマートフォン上でクリックして生み出された情報をいかに解析し、情報を抽出するかという役割を担っている<sup>13</sup>。また、膨大な量のデータの解析を確実に行うためには、リアルタイムでのデータの可視化などが必要とされるため、クラウド上ですぐに使用可能なデータへの加工なども求められている<sup>14</sup>。

データサイエンティストには、異なるソースから運ばれてくる数百万件というデータを使って消費者のパターンを読み取り、その行動を統計学のモデルにする能力が求められる。企業側は従来のマーケティング以上のものが必要と述べている。例えば、マーケットプレイスのサイトを運営する Etsy 社では、生物統計学 (Biostatistics) の博士号を持つデータサイエンティストがデータの分析を行っており、街で見かけた洋服を探すために Etsy 社のウェブサイトを検索する人がどのようなキーワードを使うか分析している。モバイルペイメントのベンチャー企業 Square 社では、認知心理学の博士号を持つデータサイエンティストが、消費者からの返金要請が多い企業に特徴的な行動パターンを調べている<sup>15</sup>。

昨年だけで約 6,000 の企業がデータサイエンティストを募集しており、LinkedIn など複数のビジネス系サイトにリストされているデータサイエンティストの求人は 24,000 件から 36,000 件となっている。しかしながら、データサイエンティストとして需要が高い統計学、生物統計学、素粒子物理学、コンピューターサイエンスの博士号の取得者は 2012 年度だけで約 2,500 人とどまっている。データサイエンティストへの需要に対応するために、コロンビア大学やバージニア大学をはじめとした 6 つの大学では、データサイエンスの修士課程を新設している<sup>16</sup>。また、優秀な人材を集める企業をサポートするために Insight Data Science Fellows Program というプログラムがスタンフォード大学と複数の技術系の企業により立ち上げられている。このプログラムでは、天体物理学、神経科学、数学などを専門とした求職者を集めており、100% の就職率となっている。博士号を取得して 2 年ほどの経験を積んだデータサイエンティストの給料は、およそ 20 万ドルから 30 万ドルとなっている<sup>17</sup>。

## (3) ビッグデータを活用するベンチャー企業

### a. 巨大な投資を受けるビッグデータ関連ベンチャー企業

ビッグデータ関連ベンチャー企業への投資は医療分野などを中心に 2013 年から急加速しており、中には近年で数社しかない数億ドル規模の投資を受けた企業もある。2008 年から 2012 年までの間にビッグデー

<sup>11</sup> <http://www.forbes.com/sites/louiscolumbus/2014/06/24/roundup-of-analytics-big-data-business-intelligence-forecasts-and-market-estimates-2014/>

<sup>12</sup> <http://www.forbes.com/sites/louiscolumbus/2014/06/24/roundup-of-analytics-big-data-business-intelligence-forecasts-and-market-estimates-2014/>

<sup>13</sup> <http://online.wsj.com/articles/academic-researchers-find-lucrative-work-as-big-data-scientists-1407543088>

<sup>14</sup> <http://www.forbes.com/sites/gilpress/2013/12/12/16-1-billion-big-data-market-2014-predictions-from-idc-and-iiia/>

<sup>15</sup> <http://online.wsj.com/articles/academic-researchers-find-lucrative-work-as-big-data-scientists-1407543088>

<sup>16</sup> <http://online.wsj.com/articles/academic-researchers-find-lucrative-work-as-big-data-scientists-1407543088>

<sup>17</sup> <http://online.wsj.com/articles/academic-researchers-find-lucrative-work-as-big-data-scientists-1407543088>

タ関連ベンチャー企業へ投資された額は合計 49 億ドルにのぼるが、2013 年は 10 月までの時点で 36 億ドルの投資が行われており、投資が急加速していることがわかる<sup>18</sup>。以下の Cloudera 社と Palantir Technologies 社はビッグデータ関連ベンチャー企業の中でも最も投資を受けている企業で、これまでに 10 億ドル前後の投資を受けており<sup>19</sup>、Dropbox 社、Airbnb 社<sup>20</sup>、Pinterest 社<sup>21</sup>などよりも大きな投資額となっている<sup>22</sup>。

2008 年に設立された Cloudera 社は、他社に先駆けて Hadoop をベースとしたビッグデータ向けプラットフォームを提供している<sup>23</sup>。これまでにベンチャーキャピタルなどから 10.4 億ドルという巨大な投資を受けており、現在ビッグデータ市場で最も大きな投資を受けているベンチャー企業となっている。この投資の中には Intel 社からの 7.4 億ドルの投資が含まれているが、同社ではこの投資がビッグデータ市場の成長を加速化させることで、新しいサーバー市場の開拓につながることを期待している<sup>24</sup>。Cloudera 社のサービスの特徴として、データベースとして整えられていない非構造化データでも、ユーザーがデータの取り扱いをできるようなシステムを提供している。近年では Hadoop で使われるデータベース向けのリアルタイムで検索が可能なくエリ(検索)エンジンをオープンソースとして公開している<sup>25</sup>。

ビッグデータの分析を行う Palantir Technologies 社は、これまでにベンチャーキャピタルなどから 9.5 億ドルの投資を受けているが<sup>26</sup>、その中には米中央情報局(Central Intelligence Agency: CIA)や国家安全保障局(National Security Agency: NSA)からの投資が含まれており、業務内容の多くは明らかにされていない<sup>27</sup>。創業者の 1 人である Peter Thiel 氏が、ロシアマフィアによる Paypal 社への詐欺行為を追跡するために開発したソフトウェアの技術が、同社のビジネスの始まりとなっており、近年では証券取引委員会(Securities and Exchange Commission: SEC)の依頼によるインサイダー取引などの違法な金融取引を見つけ出すソフトウェアの開発を行っている。同社の特徴として、データの統合を行う能力の高さや、多数の若くて優秀な人材を抱えている点が挙げられている。同社のクライアントの中には国防総省(Department of Defense: DOD)、司法省(Department of Justice)、ニューヨーク市警察(New York City Police Department)などがあるが、営業部門を持っておらず、クライアントへ非常に高額な報酬を要求することで知られている<sup>28</sup>。

## b. 急成長する医療アナリティクスサービス市場

医療分野のアナリティクス(分析)サービスに対する投資にも注目が集まっており、すでに大手ベンチャーキャピタルの多くが投資を行っている。この分野におけるベンチャー企業への投資額は 2013 年で 1 億 5,500 万ドルだったが、2014 年前半だけで 4 億ドルを超えており、年内には 10 億ドルを越すと見られている<sup>29</sup>。図表 2 は医療アナリティクスサービス市場の成長を示したものとなっている。

<sup>18</sup> <http://venturebeat.com/2013/12/09/big-data-startups-pull-in-big-money-in-2013/>

<sup>19</sup> <http://www.forbes.com/sites/gilpress/2013/10/30/top-10-most-funded-big-data-startups-updated/>

<sup>20</sup> 個人間の宿泊仲介を行うウェブサイト。190ヶ国でサービスを提供している。

<sup>21</sup> ピンボード風に表示させることができる写真投稿サイト。自由に画像コレクションを作成することができる。

<sup>22</sup> <http://www.bizjournals.com/sanjose/news/2014/08/14/meet-silicon-valleys-best-funded-tech-startups-in.html?page=all>

<sup>23</sup> <http://www.cio.com/article/2384982/big-data/10-hot-big-data-startups-to-watch.html>

<sup>24</sup> <http://www.forbes.com/sites/gilpress/2013/10/30/top-10-most-funded-big-data-startups-updated/>

<http://www.reuters.com/article/2014/03/31/us-intel-cloudera-idUSBREA2U0ME20140331>

<sup>25</sup> <http://www.cio.com/article/2384982/big-data/10-hot-big-data-startups-to-watch.html>

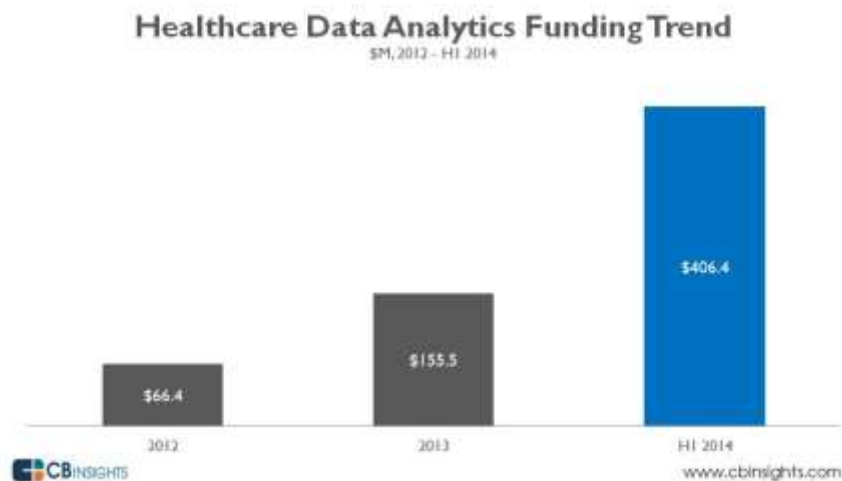
<sup>26</sup> <http://www.forbes.com/sites/gilpress/2014/10/06/top-ten-most-funded-big-data-startups-october-2014/>

<sup>27</sup> <http://www.bizjournals.com/sanjose/news/2014/09/15/cia-backed-palantir-adds-50m-to-its-war-chest.html>

<sup>28</sup> <http://www.cnbc.com/id/101487353#>

<sup>29</sup> <https://www.cbinsights.com/blog/healthcare-analytics-big-data-2014/>

図表 2: 医療アナリティクスサービス企業への投資



出典: CBINSIGHTS<sup>30</sup>

医療アナリティクスサービス市場では様々なベンチャー企業が生まれており、大手ベンチャーキャピタルはそれらの企業が設立した直後から投資を行っている。以下はベンチャーキャピタルから支援を受けている医療アナリティクスサービス関連のベンチャー企業である<sup>31</sup>。

- **Ambient Clinical Analytics 社**: 手術室や救急救命室で使用する分析ツールを開発。Mayo Clinicなどが支援を行っている。
- **Apixio 社**: 神経言語プログラミングと機械学習を使ったデータ分析ソリューションを提供。
- **Lumiata 社**: 米国の大手健康保険組合 BlueCross BlueShield に所属している医療従事者や保険会社向けに分析したデータの配信サービスを提供。
- **Healthcare Interactive 社**: 保険会社や医療関連企業向けに、病気のリスクやコストがかかりそうな患者の情報を配信するサービスを提供。
- **Zephyr Health 社**: 新しい治療法の研究開発のために必要なデータ分析ツールやアプリケーションを提供。

<sup>30</sup> <https://www.cbinsights.com/blog/healthcare-analytics-big-data-2014/>

<sup>31</sup> <https://www.cbinsights.com/blog/healthcare-analytics-big-data-2014/>

### 3 連邦政府のビッグデータ政策

#### (1) ホワイトハウスのビッグデータに関する報告書

##### a. 報告書作成の経緯

2014 年 5 月、ホワイトハウスはビッグデータとプライバシー政策の見直しを行ったものをまとめた、「ビッグデータ：機会をとらえ、価値を守る(BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES)」と題する報告書を発表した。この報告書は、同年 1 月にオバマ大統領が NSA による通話・通信記録に関する諜報活動の改革案と同時に発表された<sup>32</sup>。政府による監視活動の是正を示すと同時に、ビッグデータとプライバシーの見直しを行うことで、政府や民間企業が所有するデータのプライバシーについて政府が対応を示したものとなっている<sup>33</sup>。NSA の改革案とともに発表されたこともあり、報告書の中で NSA とビッグデータの使用について述べられると期待されたが、それについては言及されなかったことから、核心部分に触れていないという声も出ている<sup>34</sup>。図表 3 は NSA 改革案の発表を行うオバマ大統領とビッグデータ報告書を示している。

図表 3: NSA 改革案を発表するオバマ大統領とビッグデータ報告書



出典: Politico、Whitehouse<sup>35</sup>

オバマ大統領の依頼を受けた大統領顧問の John Podesta 氏が報告書のとりまとめを指揮し、ビッグデータの利点を最大限に活かしつつ、いかにリスクを最小化させるかという点に焦点が当てられている<sup>36</sup>。この報告書の作成のために、企業、研究機関、政府機関といった 113 のステークホルダーに聞き取り調査を行い、一般からの意見を集めるために 3 度のシンポジウムが開かれている<sup>37</sup>。また、この報告書と同時に大統領科学技術諮問委員会(The President's Council of Advisors on Science and Technology: PCAST) からビッグデータ政策の見直しを技術的な観点からまとめた「ビッグデータ：技術的観点(Big Data: A Technological Perspective)」が発表されている<sup>38</sup>。

<sup>32</sup> [http://www.washingtonpost.com/politics/full-text-of-president-obamas-jan-17-speech-on-nsa-reforms/2014/01/17/fa33590a-7f8c-11e3-9556-4a4bf7bcdb84\\_story.html](http://www.washingtonpost.com/politics/full-text-of-president-obamas-jan-17-speech-on-nsa-reforms/2014/01/17/fa33590a-7f8c-11e3-9556-4a4bf7bcdb84_story.html)

<sup>33</sup> <http://www.politico.com/story/2014/01/obama-vows-to-review-big-data-privacy-impact-102620.html>

<sup>34</sup> <https://www.eff.org/deeplinks/2014/05/white-house-big-data-report-good-bad-and-missing>

<sup>35</sup> <http://www.politico.com/story/2014/01/obama-vows-to-review-big-data-privacy-impact-102620.html>

[http://www.whitehouse.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_may\\_1\\_2014.pdf](http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf)

<sup>36</sup> [http://www.whitehouse.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_5.1.14\\_final\\_print.pdf](http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf)

<sup>37</sup> [http://www.whitehouse.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_5.1.14\\_final\\_print.pdf](http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf) p.69-p.73

<sup>38</sup> <http://www.whitehouse.gov/blog/2014/05/01/pcast-releases-report-big-data-and-privacy>

## b. ビッグデータ報告書による今後の課題

「ビッグデータ：機会をとらえ、価値を守る」の中では、政府や民間におけるこれまでのビッグデータの活用や取り組みなどに加え、ビッグデータと個人の関わり、公的分野や民間セクターにおけるビッグデータの取り組み、政府による政策的枠組み、政策提言などが紹介されている<sup>39</sup>。注目すべき点として、ビッグデータを最大限に活用しリスクを最小化させるために、今後取り組みを進める必要のある以下の 5 つをあげている<sup>40</sup>。

- ① **プライバシー価値の保護**：米国と、相互利用可能な国際的なプライバシーフレームワークの両方を通して、市場における個人情報を保護することでプライバシーの価値を維持する。
- ② **確実で責任のある教育**：K-12(幼稚園から高校)を中心として、学校はビッグデータを活用することで学習機会を高めることができることを認識する。同時に個人情報を保護し、デジタルリテラシーやスキルを向上させる。
- ③ **ビッグデータと差別**：ビッグデータの使用によって生み出されるかもしれない新しいタイプの差別を防ぐ。
- ④ **法の執行とセキュリティ**：法の執行、社会の安全、国家安全保障を進める中で、ビッグデータの責任ある使用を確実にする。
- ⑤ **公的資源としてのデータ**：データを公的資源として活用し、公的サービスの向上のために使用し、ビッグデータ革命を促進させる研究開発に投資を行う。

## c. 報告書の分析：7つの視点

「ビッグデータ：機会をとらえ、価値を守る」について、各界からコメントや分析が行われたが、ある専門家は、ビッグデータに関する潮流として以下の 7 つを報告書から読み取っている<sup>41</sup>。

将来性	内容
ビッグデータは確実に押しよせてきており、逆戻りできない。	ビッグデータ革命は、科学技術に関与している省庁のみならず、政府のすべての省庁、部署に深く根を下ろすことになる。特に、今まで高度なデータ分析を活発に行ってこなかった省庁にとって、ビッグデータはチャンスをもたらすことになり、市民やビジネスなどへのサービス向上に大きく役立つことが期待される。
ビッグデータは、政府全体を変革する力を持つ。	ビッグデータは連邦レベルではなく、州自治体により威力を発揮する。すでにニューヨーク市のデータ分析事業部やシカゴのスマートデータプロジェクトなど、行政サービスを向上するための例は枚挙に暇がない。
プライバシー政策は改革が必要である。	現行の電子通信プライバシー法 (ECPA) の改革は、IT 関係の業界団体 TechAmerica や the Information Technology Industry Council から大きな賛同を得ている。今まで IT 業界として、プライバシー保護に関する改革を提唱してきたが、今回その実現に向けて動きだそうとしており、心強いと言える。特に、オンラインのやり取りなどの情報のアクセスには執行を許可する裁判所の礼状を義務付けることなどのルールが必要。
個人ごとにカスタマイズされた教育の時代。	教育分野ではプライバシー保護が絶対必要である。特に子供たちに関するデータは慎重に取り扱う必要がある。それが担保できれば、ビッグデータは教育の実態に大きな変革をもたらし、個人のニーズに合った学習を提供できる。現在政府が取り組んでいる、今後 5 年間で 99% の生徒がオンラインでアクセスできるようになれば、ビッグデータの教育での活用にさらに弾みがつく。
予測分析ツールは万能ではない。	予測分析は重要な役割を果たすと考えてはいけない。例えば、法執行の分野において有罪の証明などに予測分析を使用するべきではない。また、ビッグデータを使う際には、個人のプライバシー、公民権の保護には十分注意するべきであり、憲法で定められた人権を守るために、ビッグデータの弊害も肝に銘じる必要がある。

<sup>39</sup> [http://www.whitehouse.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_5.1.14\\_final\\_print.pdf](http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf) p.v

<sup>40</sup> [http://www.whitehouse.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_5.1.14\\_final\\_print.pdf](http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf) p.59-p.60

<sup>41</sup> <http://www.govtech.com/data/7-Things-to-Know-About-the-White-House-Big-Data-Report.html>



<p>ビッグデータは国家資産である。</p>	<p>ビッグデータは米国全体の資産であることを認識しなければいけない。重要な資産であることを、広く一般の人々に利用してもらえるよう工夫すると同時に、経済発展と社会の貢献のために、しっかりと守らなければいけない。国民の資産であるという認識のもと、ビッグデータを守る、使うための重要な手段である Data.gov を整備していくことが必要となる。</p>
<p>ビッグデータの有効利用には研修や投資が必要である。</p>	<p>ビッグデータを有効に使うには、内部の研修、それに見合う予算の確保、時間など十分に用意する必要がある。今後、データの匿名化、データソースの取り調べ、暗号化、消費者が使うデータツールなどの分野で更なる研究開発が必要であり、今米国民にとって最もインパクトがある分野へ重点的な投資を行っていく必要がある。</p>

## (2) ビッグデータ研究開発イニシアチブ (Big Data Research and Development Initiative)

2012 年、オバマ政権はビッグデータの研究開発と活用を促進するための戦略として、ビッグデータ研究開発イニシアチブ (Big Data Research and Development Initiative) を発表した。このイニシアチブは、大統領科学技術諮問委員会 (The President's Council of Advisors on Science and Technology: PCAST) が大統領府に対してビッグデータへの投資が十分でないという提言に対して出されたものであり、巨大かつ複雑なデータから様々な知識や物事の本質を引き出す能力を高めることで、国家が解決すべき課題への対応に役立てることを目的としている。6 つの連邦政府機関に対してビッグデータの活用と研究開発を行うための資金として 2 億ドルが投資されている。以下は研究開発が行われている 6 つの政府機関である<sup>42</sup>。

- 国立科学財団 (National Science Foundation: NSF)
- 国立衛生研究所 (National Institutes of Health: NIH)
- 国防総省 (Department of Defense: DOD)
- エネルギー省 (Department of Energy: DOE)
- 国防高等研究計画局 (Defense Advanced Research Projects Agency: DARPA)
- 米地質調査所 (United States Geological Survey: USGS)

この研究開発イニシアチブの中では、NIH と Amazon Web Services 社による 200 テラバイトの遺伝子データを研究者向けに公開する「1000 Genomes Project」や、巨大なデータから知覚、認知、意思決定を行い自律システムの研究へと役立てる DOD のプロジェクトなどの研究プログラムが紹介されている<sup>43</sup>。2013 年 11 月には、大統領府の科学技術政策局 (Office of Science and Technology Policy: OSTP) とネットワーク及び情報技術研究開発 (Networking and Information Technology Research and Development: NITRD) から、新しい産学官のビッグデータパートナーシップが発表された。これは前述のビッグデータ研究開発イニシアチブから生まれたものであり、連邦省庁、州政府、地方自治体と、企業、教育機関、専門組織が連携して行う、新たなビッグデータの共同事業となっている<sup>44</sup>。以下に研究開発イニシアチブの一部を紹介する。

<sup>42</sup> [http://www.whitehouse.gov/sites/default/files/microsites/ostp/big\\_data\\_press\\_release\\_final\\_2.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf)

<sup>43</sup> [http://www.whitehouse.gov/sites/default/files/microsites/ostp/big\\_data\\_press\\_release\\_final\\_2.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf)

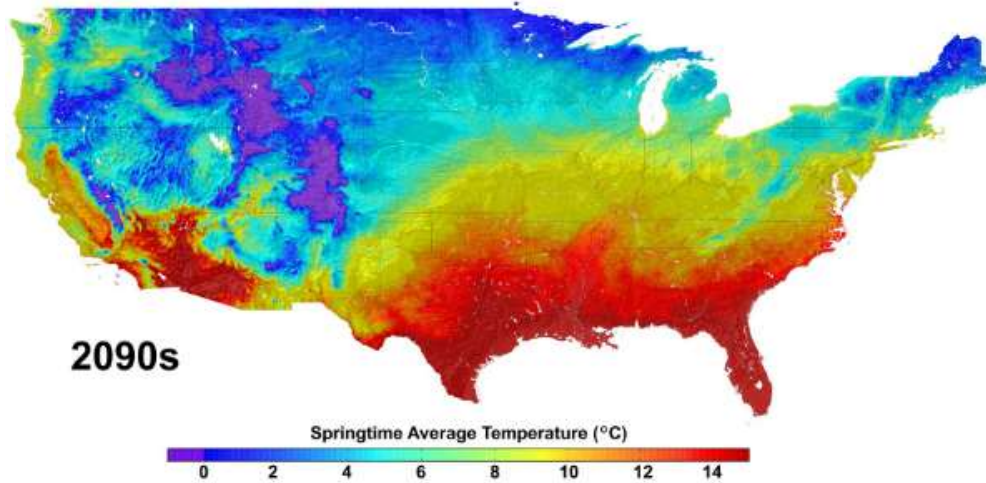
<sup>44</sup> <http://www.informationweek.com/big-data/big-data-analytics/white-house-unveils-big-data-projects-round-two/d/d-id/1112226>

<http://www.whitehouse.gov/sites/default/files/microsites/ostp/Data2Action%20Announcements.pdf>

**a. NASA と Amazon 社による地球観測データの公開**

航空宇宙局(National Aeronautics and Space Administration : NASA)は、Amazon Web Services (AWS)社と協力して地球観測データを研究者向けに公開するNASA Earth eXchange (NEX)プログラムを進めており、NASAが保有する膨大な観測データをAWS社のクラウドサービスを通して一般に公開している。NEXでは、地球科学のデータセットの分析、アルゴリズムモデルの共有、科学プロジェクトの管理、研究コミュニティ間での協力などができるようになっている。現在、AWS社を通して公開しているデータとして、米国本土の高解像度の天候変動予測ができる小型気候シミュレーション、中解像度画像分光放射計(Moderate Resolution Imaging Spectroradiometer : MODIS)から送られる地球の観測映像、地質調査所(United States Geological Survey : USGS)から送られる地球観測衛星(LANDSAT)の地表データなどがある<sup>45</sup>。図表4はNASA NEXプログラムの気候シミュレーションのイメージとなっている。

**図表 4: NASA NEX、気候シミュレーション**



出典: NASA<sup>46</sup>

**b. 産業界のデータサイエンティストを結ぶ試み**

非営利団体の DataKind は、産業界のデータ分析サービスを提供する Pivotal 社と協力して、データを所有しては活用できていない社会団体にデータサイエンティストをボランティアで派遣する試みを進めている。DataKind では、DataCorps と呼ばれる、人々に奉仕を行うデータサイエンティストのグループを派遣しており、政府機関や NGO と協力してデータの整備、分析、可視化などを行っている。最近では、十分な医療を受けられない地域へ支援を行う非営利団体 Medic Mobile と協力して、様々な医療政策の効果が行き届いているかどうか数値上で調査する取り組みを行っている。この取り組みからは、携帯電話を使った遠隔治療などを行うことで、世界規模で数百万人の健康改善の成功につながっている<sup>47</sup>。

<sup>45</sup> <http://www.whitehouse.gov/sites/default/files/microsites/ostp/Data2Action%20Announcements.pdf>

<http://aws.amazon.com/nasa/nex/>

<sup>46</sup> <http://www.nasa.gov/press/2013/november/nasa-brings-earth-science-big-data-to-the-cloud-with-amazon-web-services/#.VEDI4fI5MTY>

<sup>47</sup> <http://www.whitehouse.gov/sites/default/files/microsites/ostp/Data2Action%20Announcements.pdf>

<http://www.datakind.org/howitworks/datacorps/>

## 4 ビッグデータの活用

### (1) 医療

#### a. ウェアラブル端末とビッグデータの活用

Intel 社では、パーキンソン病の研究にウェアラブル端末とビッグデータを使った取り組みを行っている。同社はパーキンソン病の研究に取り組むマイケル・J・フォックス財団 (Michael J. Fox Foundation: MJFF) と協力してパーキンソン病の研究を行っており、ウェアラブル端末を通じた情報を分析することで患者の治療や新しい治療方法に役立てることを目的としている。ウェアラブル端末の使用により、研究者がより多くの患者を調べることができるだけでなく、遠隔地の患者にも参加してもらうことが可能となった<sup>48</sup>。

最初の段階として、9 人のボランティアと 16 人の患者にウェアラブル端末を着けてもらい、4 日間にわたってデータを集める。集められた全員のデータは匿名となっており、データの転送や分析においても情報はセキュリティで守られている。1 つの端末で 1 分間に 300 のデータを集めることができ、Intel 社のビッグデータプラットフォーム上でデータの分析を行い、リアルタイムで患者の変化が分かるようになっている。研究者は客観的に病気の進行を診断することができ、患者にどのような治療が必要か判断できるようになる。将来的には患者の遺伝子情報や臨床データと機械学習などを併用することで研究者の大きな助けになることが期待されている<sup>49</sup>。図表 5 はウェアラブル端末とビッグデータを使ったソリューションのイメージとなっている。

図表 5: ウェアラブル端末とビッグデータを使ったソリューション



出典: Intel<sup>50</sup>

#### b. IBM ワトソンを使ったデータ分析

IBM 社が開発した人工知能ワトソン (Watson) をビッグデータの分析に使う試みが行われており、医療分野で取り組みが始まっている。バージニア州の Carilion Clinic では、ワトソンの自然言語処理の技術とデータ分析の能力を使い、特定の病気のリスクを持つ患者を探す取り組みを行っている<sup>51</sup>。心臓病は入院患者がもっとも多い病気の 1 つであり、電子カルテを使用しても、詳細な情報が医療メモに埋もれてしまうことも少なくないため、Carilion Clinic は IBM 社と協力して医療データから心臓病患者を予測分析する取り組みを行っている。

<sup>48</sup> <http://www.businessinsider.com/r-intel-explores-wearable-devices-for-parkinsons-disease-research-2014-13>

<sup>49</sup> <http://www.informationweek.com/healthcare/analytics/intel-michael-j-fox-foundation-take-on-parkinsons-disease/d/d-id/1298010>

<sup>50</sup> [http://newsroom.intel.com/community/intel\\_newsroom/blog/2014/08/13/the-michael-j-fox-foundation-and-intel-join-forces-to-improve-parkinsons-disease-monitoring-and-treatment-through-advanced-technologies](http://newsroom.intel.com/community/intel_newsroom/blog/2014/08/13/the-michael-j-fox-foundation-and-intel-join-forces-to-improve-parkinsons-disease-monitoring-and-treatment-through-advanced-technologies)

<sup>51</sup> <http://www.forbes.com/sites/zinamoukheiber/2014/02/19/ibm-and-epic-apply-predictive-analytics-to-electronic-health-records/>

35 万人の患者の情報の他に、医療記録の欄には書かれていない 2,000 万件分のメモなども使用している。体系化されていない医療メモや用済みとなった書類などの分析に自然言語を処理する技術を使うことで、各患者の詳細な情報がわかっている<sup>52</sup>。集めたデータを基に予測分析を行った結果、1 年以内に心臓病を発症する可能性がある患者 8,500 人の特定に成功しており、85%の精度となっている。その中の 3,500 人は従来の分析では見落としていた可能性がある<sup>53</sup>と見ている。

## (2) 教育

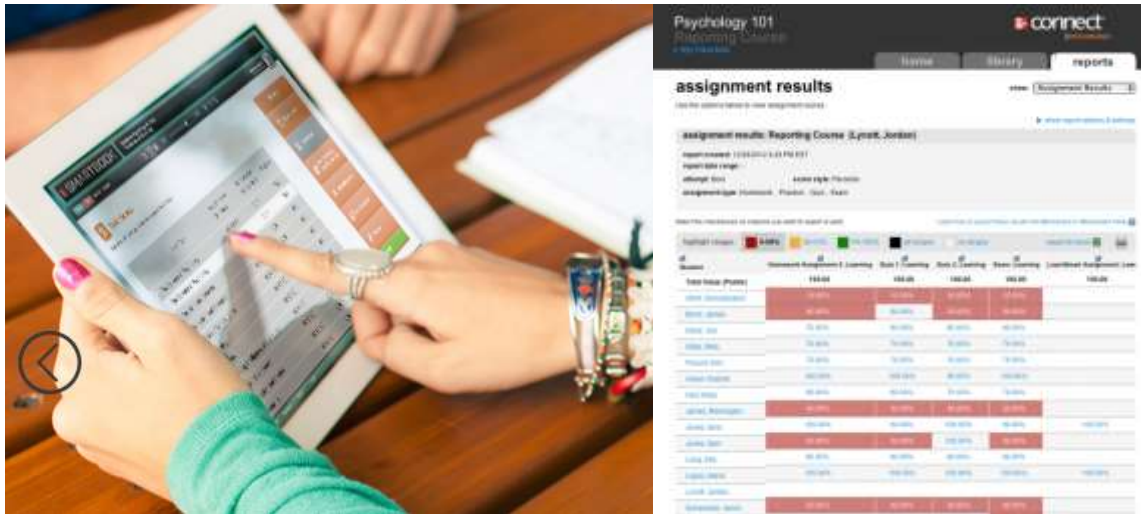
### a. 多様なビッグデータの活用

米国では、K-12(幼稚園から高校)を中心に、生徒や学校に関する情報を活用する動きが急速に広まっている。教師が州政府のウェブサイトを通して生徒の情報にアクセスすることができるサービスは、2011 年に 28 州でしか提供されていなかったが、現在では 35 州にまで拡大している。その中の 17 州では教師のトレーニングプログラムに生徒の情報を活用しており、14 の州では保護者が子供の情報にアクセスできるようになっている<sup>54</sup>。この他、以下のような様々な形でビッグデータを活用する取り組みが進んでいる。

- 生徒ごとに最適な学習を提供する「適応学習(Adaptive learning)」

McGraw-Hill Education 社が提供する学習プログラムでは、人工知能を使って 200 万人の学生のデータ分析を行い、各生徒に適した学習カリキュラムとコースを提供している。質問の回答によって適した教材を提供し、どの問題に回答可能かシステムが判断することで、生徒の集中力が続く学習システムを作り出している<sup>55</sup>。図表 6 は適応学習の製品のイメージとなっている。

図表 6: 適応学習の製品のイメージ



出典: McGraw-Hill Education<sup>56</sup>

<sup>52</sup> <http://www-03.ibm.com/press/us/en/pressrelease/43231.wss>

<http://www.forbes.com/sites/zinamoukheiber/2014/02/19/ibm-and-epic-apply-predictive-analytics-to-electronic-health-records/>

<sup>53</sup> <http://www-03.ibm.com/press/us/en/pressrelease/43231.wss>

<http://www.forbes.com/sites/zinamoukheiber/2014/02/19/ibm-and-epic-apply-predictive-analytics-to-electronic-health-records/>

<sup>54</sup> [http://www.huffingtonpost.com/2013/11/19/schools-collecting-student-data\\_n\\_4302346.html](http://www.huffingtonpost.com/2013/11/19/schools-collecting-student-data_n_4302346.html)

<sup>55</sup> <http://www.forbes.com/sites/emc/2014/03/06/big-data-goes-to-school/2/>

<http://campustechnology.com/articles/2014/03/03/mcgraw-hill-education-intros-real-time-analytics-on-student-performance.aspx>

<sup>56</sup> <http://connect.customer.mheducation.com/features/instructors/#easy-to-use?feature=simple-lms-integration>

- 教育現場への寄付システム

非営利団体 DonorsChoose では、公立学校の教師がクラスで取り組みたいプロジェクトに対して寄付を集めるクラウドファンディングサイトを運営している。投稿されるプロジェクトはメールやソーシャルネットワーキングを通して受け付けており、100 万件を超えるプロジェクトのうち約 80%は達成につながっている。DonorsChoose は、これまでのプロジェクトやトレンドを一般に公開しており、データを分析することで、学校長や運営機関などに対して、現場の教師がどのようなものを求めているか知る手がかりになると語っている。データは要望に応じて、地区、貧富、人種などのカテゴリ分けすることが可能となっている<sup>57</sup>。

- 能力の高い教師の雇用への活用

カリフォルニア州 Chula Vista をはじめとした複数の都市では、教師の雇用にビッグデータを活用している。これらの都市は、マーケティング企業 Hanover Research 社が行っている Paragon K-12 と呼ばれる、予測分析ソフトウェアの使用を行うパイロットプログラムに選ばれており、選ばれた学校は教師の履歴書から、学歴、哲学、態度、経験、信念といった複数の要因を、ソフトウェアを使って分析し、雇用の検討を行っている。

## b. inBloom

米国では、教育レベルを向上させるためにビッグデータを活用した様々な取り組みが展開されているが、inBloom のように多額の投資を受けながら、生徒のプライバシー保護に対する過剰な反応により頓挫したケースがある。ジョージア州アトランタに拠点を置く非営利団体 inBloom は、地区、州政府、教育関連企業と連携して公立学校の生徒に関するデータの管理を一括して行うサービスの提供を目指し、2011 年に設立された。同組織のデータ管理システムにはセキュリティが高いものを用意し、データの取り扱いについても明確な説明を行うなど、公共性の高さやデータの取り扱いに関する透明性が認められ、ビル・アンド・メリンダ・ゲイツ財団 (Bill & Melinda Gates Foundation) やカーネギー財団 (Carnegie Foundation of New York) などから合計 1 億ドルの出資を受けるなど、大きな期待が寄せられていた。

inBloom はサービス開始後 9 州の教育機関と契約を結び、集めたデータの分析を行って、教師や保護者が分析結果にアクセスできることを約束していた。しかしながら、保護者の一部やプライバシー保護団体などは、子供の情報が売買される恐れやハッカーによって情報が盗まれる懸念<sup>58</sup>、情報の共有について許可を求められていないことや、データのセキュリティに完全な保証が出されていない点について inBloom を非難する人もいた。また、生徒の情報にはランダムに番号が割り振られていたものの、収集された情報の中に社会保障番号 (Social Security number) が入っていたことが指摘されていた。さらに、非営利団体電子プライバシー情報センター (Electronic Privacy Information Center) が、inBloom のデータベースが生徒のプライバシーに深刻な脅威をもたらすもとして、教育省 (Department of Education) への告訴に踏み切っている<sup>59</sup>。こうした声が高まる中で、契約を結んでいた 9 州のうち 6 州がプライバシーの懸念から離脱し、2014 年 4 月 21 日に inBloom は閉鎖することを発表した<sup>60</sup>。

---

<http://learnsmartadvantage.com/>

<sup>57</sup> <http://www.forbes.com/sites/emc/2014/03/06/big-data-goes-to-school/2/>

<sup>58</sup> <http://www.businessweek.com/articles/2014-05-01/inbloom-shuts-down-amid-privacy-fears-over-student-data-tracking>

<sup>59</sup> <http://www.washingtonpost.com/blogs/answer-sheet/wp/2013/06/09/privacy-concerns-grow-over-gates-funded-student-database/>

<sup>60</sup> <http://www.businessweek.com/articles/2014-05-01/inbloom-shuts-down-amid-privacy-fears-over-student-data-tracking>

### (3) 社会サービス

#### a. 子供の虐待防止への活用

フロリダ州の児童家族局 (Department of Children and Families: DCF) は、虐待によって命を落とす子供の現状を把握し、虐待につながる要因を特定するためにビッグデータを使った調査を行っている。この調査では、約 100 万人分の子どもの情報を基にした分析を行っており、虐待の傾向を探るだけでなく、将来的には予測分析を使った子供への福祉サービスなども視野に入れている。ビッグデータの分析には、コンサルティング企業の North Highland 社、データ分析会社の SAS Institute 社、非営利組織 Child Welfare Policy and Practice Group などと協力して進めている<sup>61</sup>。

分析したデータは、Florida Safe Families Network<sup>62</sup>や、フロリダ州の子供の死に関する情報を集めた Child Death Review Database などの情報を中心に分析が行われており、2007 年 1 月 1 日から 2013 年 6 月 30 日までの間にフロリダ虐待ホットラインに報告された情報なども使われている。データの使用に際しては、個人識別が可能な情報を取り除いた上で、地域、社会サービス、家族関係、虐待の疑い、保護の状況など様々な要因を分析している<sup>63</sup>。

5 週間にわたる調査からは、全体的に子供の死は減少傾向にあるが、DCF が前もって家庭訪問することにより子供の虐待死の危険性が 90%減少することがわかり、ケースワーカーによる家庭訪問が非常に良い効果をあげていることが判明した。また、子供が以前に身体的な虐待を受けている子供は死の危険性が 14 倍増加し、両親にアルコールやドラッグなどの中毒がある場合は死の危険性が 15 倍増加することがわかっている<sup>64</sup>。プロジェクトからは、今後の児童データ活用のためにデータ整備について以下のような提言が出されている<sup>65</sup>。

1. **信頼できるデータの入力:** 自由形式のテキストデータよりも、一貫した選択肢を入力できる項目を増やす。
2. **不明瞭な入力項目、選択肢:** わかりにくい入力項目があるとデータに誤りが出てくるため、項目名や選択肢を修正する。
3. **追跡履歴:** 子どもの情報を充実させるために、家族の構成、調査員の業務など重要な情報は追跡できるようにする。
4. **トレーニングや書類のレビュー:** データの正確性や整合性を確保するために、データ入力のトレーニングや入力内容の確認を行う。
5. **地域サービスからのケアレポート:** 地域のケースワーカーからのレポートを改善させることで、より確固としたデータが得られるようにする。

#### b. ホームレス問題への取り組み

ホームレス支援団体とシカゴ大学は、ビッグデータを使ってホームレス問題への取り組みを行っている。複数のホームレス支援団体で構成される Chicago Alliance と、シカゴ大学 (University of Chicago) のデータサイエンスを社会貢献に役立てる活動を行っている Data Science for Social Good (DSSG) は、ホームレス問題への取り組みのためにビッグデータの分析を 2013 年から行っている<sup>66</sup>。

<sup>61</sup> [http://www.dcf.state.fl.us/newsroom/pressreleases/docs/20140106\\_pressrelease\\_attachments.pdf](http://www.dcf.state.fl.us/newsroom/pressreleases/docs/20140106_pressrelease_attachments.pdf) p.1

[http://www.sas.com/en\\_us/news/press-releases/2014/july/predictive-analytics-government.html](http://www.sas.com/en_us/news/press-releases/2014/july/predictive-analytics-government.html)

<sup>62</sup> フロリダ州の児童福祉情報に関するシステム。

<sup>63</sup> [http://www.dcf.state.fl.us/newsroom/pressreleases/docs/20140106\\_pressrelease\\_attachments.pdf](http://www.dcf.state.fl.us/newsroom/pressreleases/docs/20140106_pressrelease_attachments.pdf)

<sup>64</sup> [http://www.dcf.state.fl.us/newsroom/pressreleases/docs/20140106\\_pressrelease\\_attachments.pdf](http://www.dcf.state.fl.us/newsroom/pressreleases/docs/20140106_pressrelease_attachments.pdf)

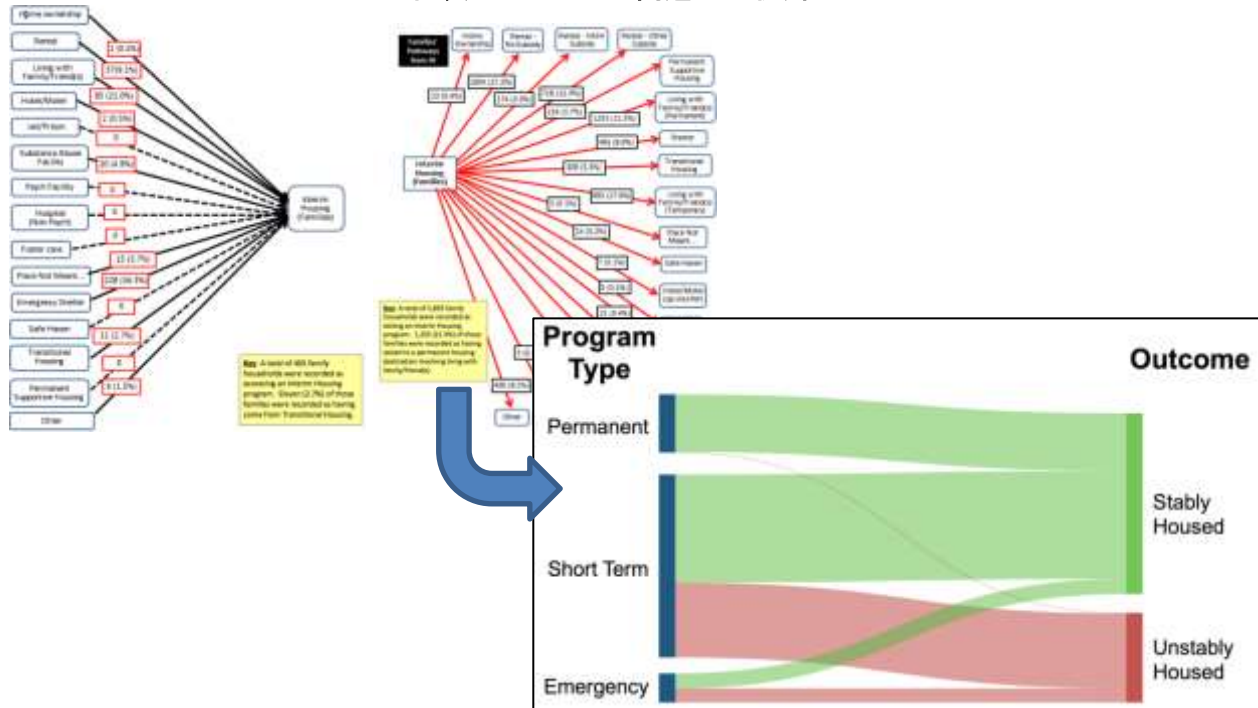
<sup>65</sup> [http://www.dcf.state.fl.us/newsroom/pressreleases/docs/20140106\\_pressrelease\\_attachments.pdf](http://www.dcf.state.fl.us/newsroom/pressreleases/docs/20140106_pressrelease_attachments.pdf)

<sup>66</sup> <http://dssg.io/2014/08/20/paths-homelessness.html>

<http://www.allchicago.org/division/chicago-alliance>

DSSG には Chicago Alliance から提供された 15 年分の匿名化されたホームレスや地域のデータが提供されている。しかしながら、これらのデータの多くは内容が抜けているものや矛盾したものが多かったため、データ整備が行われた最近 2 年のデータに限定した分析を行っている。また、使用できるデータ項目も年齢、サポートプログラム、以前の生活など限られたものとなっていたため、どこまで分析が可能かわからないという懸念があった。このため、プロジェクトではデータの視覚化を行うことで、異なる種類のサポートがどのような効果を出すかという部分に焦点を当てている。図表 7 の左は初期の視覚化したモデルとなっており、どのようなサービスの利用が仮住居への移動につながり、仮住居から安定した生活へつながるにはどのようなサービスが必要か示している。この情報から作り出した右の図からはより長期的なサービスほどホームレスの安定した生活につながることがわかっている<sup>67</sup>。

図表 7: ホームレス問題への取り組み



出典: Data Science for Social Good<sup>68</sup>

<sup>67</sup> <http://dssg.io/2014/08/20/paths-homelessness.html>

<sup>68</sup> <http://dssg.io/2014/08/20/paths-homelessness.html>

## 5 プライバシー問題

### (1) データブローカーを規制する動き

日々生み出される情報の収集を行うデータブローカーに対して、連邦取引委員会 (Federal Trade Commission: FTC) は、消費者のプライバシー保護のために早急な法規制の確立に向けて動いている。データブローカーとは、消費者の個人情報を集め、マーケティングなどの目的で収集した情報を企業へ販売する事業者を指しており、ビッグデータの経済の中で重要な役割を持っている<sup>69</sup>。データブローカーは膨大な米国消費者の個人情報を取り扱っており、その量は 14 億人分の取引記録と 7,000 億件のデータを所有し、毎月 30 億件のデータが追加されていると言われている。2014 年 5 月に FTC より発表されたデータブローカーについて調査した報告書、「データブローカー: 透明性と説明責任を求める (Data Brokers - A Call for Transparency and Accountability)」では、以下のデータブローカー 9 企業に対して調査を行っている<sup>70</sup>。

企業名	企業概要
Acxiom 社	消費者の情報や、マーケティングや不正行為検知のために分析したデータを提供。7 億人分の情報の他に、3,000 に分類されたほぼすべての米国消費者の情報を保有している。
Corelogic 社	土地資産、金融、消費者の情報の他に、分析サービスを企業や政府などへ提供している。データベースには約 8 億件の土地の売買情報、9,300 万件のローン申込書などがあり、土地に関する 1 億 4,700 万件の情報は米国の住宅の 99% をカバーしている。
Datalogix 社	米国におけるほとんどの世帯のマーケティング情報と、1 兆ドルにおよぶ取引情報を提供している。2012 年に Facebook とパートナーシップを結び、Facebook 上での広告の閲覧と、店舗での売り上げの比較を行っている。
eBureau 社	マーケティング、金融、オンラインストアなどに対し、予測分析した情報を提供しており、どの消費者が利益につながるか、詐欺行為を行う可能性があるか、といった内容となっている。数十億件分の消費者の情報に基づいた情報となっており、毎月 300 万件の情報が追加されている。
ID Analytics 社	個人の識別や、取引に詐欺行為が含まれていないか確認するサービスを提供している。同社のネットワーク内には、数百億件の様々なデータ、11 億人分の個人識別が可能な情報、14 億件の取引情報などがある。
Intelius 社	消費者の身辺情報や公的情報を、企業と消費者へ提供している。200 億件以上の情報を保有している。
PeekYou 社	60 以上のソーシャルメディア、ニュース、ウェブサイト、ブログなどを分析することができる技術を使って、顧客へ消費者の情報を提供している。
Rapleaf 社	米国消費者 80% のメールアドレスが登録されているデータベースを保有しており、メールアドレスに年齢、性別、配偶者の有無といった 30 以上の属性が関連付けられている。
Recorded Future 社	インターネット上にある消費者や企業の様々な情報を分析して、今後どのような行動を起こすか予測するサービスを提供している。2014 年 5 月までに 50 万以上のウェブサイトアクセスしている。

上記 9 社の企業概要からわかるように、収集した情報を提供するだけでなく、マーケティング、ID 識別や不正行為検知などのリスク軽減 (Risk Mitigation)、個人情報検索など、高度な分析を行ったサービスを提供している。このデータブローカー 9 社の 2012 年の利益だけで 4 億 2,600 万ドルに上っており、データサ

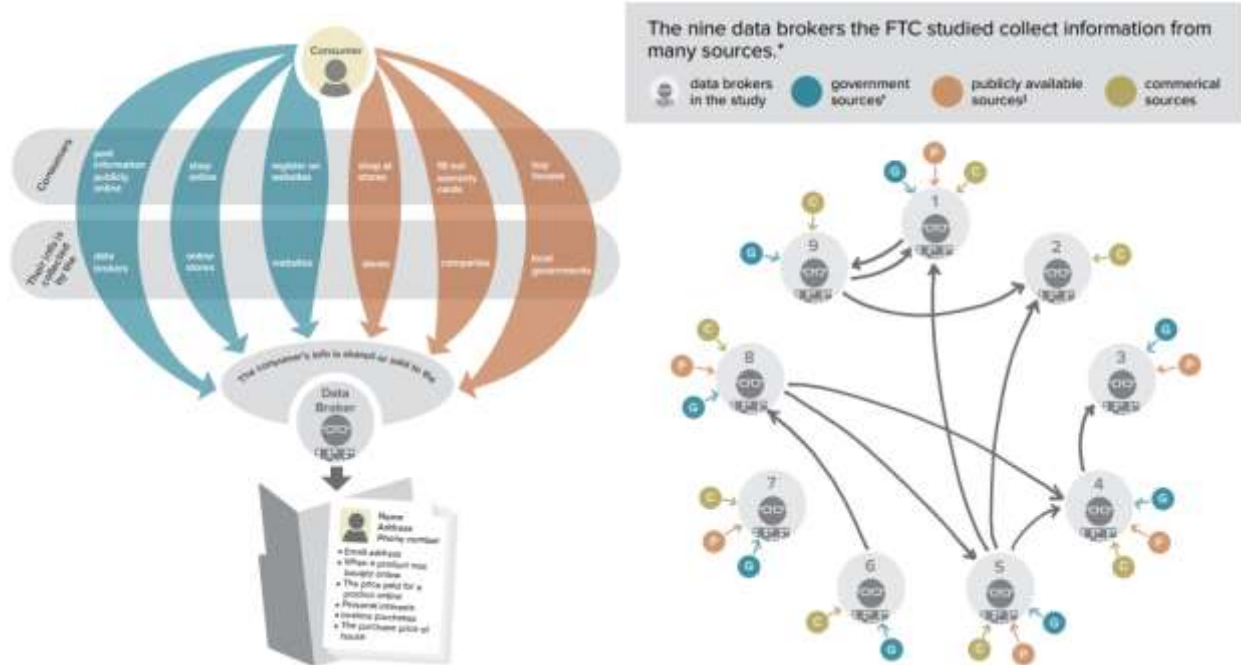
<sup>69</sup> <http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf>

<sup>70</sup> <http://www.ftc.gov/news-events/press-releases/2014/05/ftc-recommends-congress-require-data-broker-industry-be-more>



ービスは非常に大きな市場となっている<sup>71</sup>。これらの情報は、携帯電話、家や車の購入、雑誌の購読、アンケートの回答やソーシャルメディアなど、個人の日常生活における様々な場面で発生しているものであり、データブローカーはこれらの情報を収集して個人のプロファイルを作り出している。また、上記のデータブローカー9社のうち7社はお互いに集めた情報を売買しており、消費者は自身の個人情報はどこから出てきたものか事実上判別できなくなっている<sup>72</sup>。図表8は、データブローカーが行っている情報収集のイメージとなっている。

図表8: データブローカーによる情報収集のイメージ



出典: Data Brokers – A Call for Transparency and Accountability<sup>73</sup>

FTC の報告書では、自分自身の個人情報がかどのように取り扱われているかわからず、個人の承諾無しに個人情報が売買されている事態を指摘しており、データブローカー産業において透明性 (transparency) が無いことが根本にあると述べられている。このため、消費者のプライバシー保護のために次の4点を盛り込んだ、データブローカーを規制する法律の立法化を提言している<sup>74</sup>。

1. 消費者が、データブローカーが個人情報を保有していることを容易に知ることができるようにし、その情報に対してオプトアウト<sup>75</sup>する権利を行使できるようにする。
2. データブローカーに対し、名前、住所、年齢、収入などの収集した元情報を使用するだけでなく、その情報から派生したデータを作り出すことについて、ウェブサイトなどで明確に説明させる。
3. データブローカーに対し、消費者が個人情報の出所がどこなのか判断しやすくするために、データソースの名前やカテゴリを開示させる。

<sup>71</sup> <http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf> p.23

<sup>72</sup> <http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf> p.1, p.14

<sup>73</sup> <http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf>

<sup>74</sup> <http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf> p.viii

<sup>75</sup> 自身の情報をマーケティングなどに使われないようにする手段を指す。

4. 個人情報データブローカーに共有されていることを消費者に対して明確に知らせることができるシステムや、消費者が自身の情報の使用についてオプトアウトなどの選択肢を与えるシステムを構築する。

2014 年 5 月に出されたビッグデータの報告書、「ビッグデータ:機会をとらえ、価値を守る」の中でも、消費者が正しくない情報に対して修正や差し止めができる取り組みを進める必要があると述べられており、データブローカーから消費者に対して、自身の個人情報にアクセスする手段が用意されていないことがわかる<sup>76</sup>。FTC から報告書が発表される前の 2014 年 2 月には、Jay Rockefeller 上院議員(民主党、ウェストバージニア州)からデータブローカーを規制する法案が出されているが、これに対して Acxiom 社などが加入しているダイレクトマーケティング協会(Direct Marketing Association)は反発している<sup>77</sup>。

SSNDOB(現在は閉鎖)というアンダーグラウンドサイトでは、様々な個人情報やクレジットカード情報が 1 件あたり 50 セントから 2.5 ドルで売買されていた。その中には歌手・ハリウッドスターの Beyonce、大統領夫人の Michelle Obama 氏、当時の FBI 長官の情報も含まれていたが、これらの情報がどこから仕入れられたものかは謎に包まれていた。2013 年 9 月、セキュリティ専門家の Brian Krebs 氏<sup>78</sup>が発表した内容によると、SSNDOB はボットネット<sup>79</sup>を使って米国内にある複数の大手データブローカーのサーバーに侵入した形跡が見られ、この中には大手リサーチデータベース企業 LexisNexis 社や信用調査会社の Dun & Bradstreet 社などが含まれていた。LexisNexis 社はサーバーへの侵入については認めているものの、消費者の情報が盗まれた形跡はないと語っている<sup>80</sup>。

## (2) 州政府による取り組み

### a. カリフォルニア州

カリフォルニア州は全米でもプライバシー問題に積極的に取り組んできた州の 1 つであり、インターネットに関連したプライバシー保護に特化した取り組みを続けている。2003 年に出されたカリフォルニアオンラインプライバシー保護法(California Online Privacy Protection Act: CalOPPA)は、個人識別が可能な情報(Personally Identifiable Information)を収集する商業サイトの事業者に対して、プライバシーポリシーの明確な記載を定めた全米で初めての法律となっている。主に収集される情報の種類や、他の事業者と共有される可能性、ユーザーからの確認や変更の要求に対する方法などの記載を求めている<sup>81</sup>。2013 年 9 月の CalOPPA の改正では、ウェブサイト事業者が、ウェブブラウザを通して送られる Do Not Track 信号や類似した機能にどのように対応するかや、他の事業者が個人識別が可能な情報を収集しているかどうかについて、プライバシーポリシーに記載することが追加されている<sup>82</sup>。2014 年 5 月にカリフォルニア州司法長官府(Office of the Attorney General, California)から出されたプライバシーポリシーのガイドライン「Making Your Privacy Practices Public」の中でも、Do Not Track 機能について同様の観点からプライバシーポリシーを明記することなどが記されている<sup>83</sup>。

2013 年 9 月カリフォルニア州は、18 歳未満の未成年者に限りインターネット上の情報を消去する権利を認める法案を可決した。この法律では、ウェブサイトやモバイルアプリケーションなどを提供する事業者に対し

<sup>76</sup> [http://www.whitehouse.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_may\\_1\\_2014.pdf](http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf) p.47

<sup>77</sup> <http://thedma.org/news/dma-expresses-disappointment-with-new-data-broker-bill/>

<sup>78</sup> Target 社の情報漏洩について最初に発表したことなどで有名。

<http://www.computerworld.com/article/2487425/cybercrime-hacking/target-breach-happened-because-of-a-basic-network-segmentation-error.html>

<sup>79</sup> 悪意のあるプログラムで乗っ取ったコンピューターと、それらで構成されるネットワーク

<sup>80</sup> <http://krebsonsecurity.com/2013/09/data-broker-giants-hacked-by-id-theft-service/>

<sup>81</sup> <http://www.forbes.com/sites/gregorymcneal/2014/05/21/california-ag-releases-guide-to-californias-online-privacy-laws/>

<http://www.leginfo.ca.gov/cgi-bin/displaycode?section=bpc&group=22001-23000&file=22575-22579>

<sup>82</sup> [http://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill\\_id=201320140AB370](http://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201320140AB370)

<sup>83</sup> [https://oag.ca.gov/sites/all/files/agweb/pdfs/cybersecurity/making\\_your\\_privacy\\_practices\\_public.pdf](https://oag.ca.gov/sites/all/files/agweb/pdfs/cybersecurity/making_your_privacy_practices_public.pdf) p.2

て、未成年者を対象としたサービスを提供している場合や利用者が未成年とわかっている場合には、ユーザー自身が投稿した情報を削除できる機能を用意し、その方法について説明を行うことを義務付けるものとなっている<sup>84</sup>。しかしながら、この法律が現実的でないという声が上がっている。例えば、ユーザーは自分が投稿した情報しか削除できないため、他人が投稿した画像などはこの法律は適用されない<sup>85</sup>。また、削除の要請はユーザーが未成年の間しかできないため、小さい子供に削除の判断ができないのでは、といった懸念が出ている。この法案は 2015 年 1 月 1 日から施行の予定となっている<sup>86</sup>。

## b. フロリダ州

フロリダ州は、米国のすべての企業や組織に対し、フロリダに住所を持つ企業や個人に関する個人情報の流出があった場合に 30 日以内にフロリダ州法務局と消費者に対して連絡することを義務付ける Florida Information Protection Act を 2014 年 7 月に施行した。従来の 45 日から短縮した期限に設定しており、個人情報の流出が起きた場合に、より短期間で情報保護を行い、情報流出に迅速に対応することを目的としている<sup>87</sup>。また、個人情報の定義も拡大されており、オンライン上の個人のアカウントにアクセスするために必要なユーザーネームやメールアドレスの他に、パスワードやセキュリティ質問の回答などの組み合わせ個人情報と見なすようになっている<sup>88</sup>。個人情報の中にオンラインアカウントを含めた法律を制定している州にはカリフォルニア州などがあり、全米でも数少ない取り組みの 1 つとなっている<sup>89</sup>。

## (3) 米国における「忘れられる権利」

2014 年欧州司法裁判所は、インターネット関連企業に対して、無関係または行き過ぎた個人情報を検索結果から削除することを命じる判決を出した。これは 1998 年に新聞に掲載された住宅の差し押さえに関する記事が、スペイン人の男性の名前入りで Google 社の検索結果から出てくることに抗議していた問題に対する判決となっており、インターネット上の「忘れられる権利(right to be forgotten)」を認めた形となっている。この判決を受けて Google 社は、EU 圏での削除要請を受け付けるようになっている。これまでに 14 万 4,954 件の削除要請を受けているが、そのうちの 58%の削除要請を拒否している<sup>90</sup>。

米国国内では現在そのような権利は認められておらず、EU での判決に対して New York Times 紙は、報道の自由と言論・表現の自由に反すると抗議している。米国では歴史背景や法制度が違うため、このような事例はなりたたないとされているが<sup>91</sup>、米国の Google 社も「忘れられる権利」を適応すべきかとの議論が多く繰り広げられており、ネット上のプライバシー情報についての議論にも影響を与えている<sup>92</sup>。

Pew Research Center の調査によると、米国において 68%のインターネット利用者がオンライン上の個人情報を守るのに米国法律は十分ではないと感じることが判明している。しかしながら米国では、米国憲

<sup>84</sup> [http://leginfo.ca.gov/faces/billNavClient.xhtml?bill\\_id=201320140SB568](http://leginfo.ca.gov/faces/billNavClient.xhtml?bill_id=201320140SB568)

<sup>85</sup> <http://techcrunch.com/2013/09/24/on-californias-bizarre-internet-eraser-law-for-teenagers/>

<sup>86</sup> <http://www.forbes.com/sites/ericgoldman/2013/09/24/californias-new-online-eraser-law-should-be-erased/>

<sup>87</sup> <http://www.bizjournals.com/tampabay/blog/morning-edition/2014/07/florida-information-protection-act-shines-light-on.html>

<http://www.mondaq.com/unitedstates/x/327078/data+protection/New+Florida+Information+Protection+Act+Expands+Data+Breach+Notification+Requirements>

<sup>88</sup> <http://www.businessinsurance.com/article/20140624/NEWS07/140629916?tags=%7C299%7C69%7C329%7C305%7C80%7C303%7C87>

<sup>89</sup> <http://www.scmagazine.com/california-law-would-require-breach-notice-if-online-account-information-is-stolen/article/294296/>

<http://www.businessinsurance.com/article/20140624/NEWS07/140629916?tags=%7C299%7C69%7C329%7C305%7C80%7C303%7C87>

<sup>90</sup> <http://www.newsweek.com/how-spaniards-eviction-created-right-be-forgotten-google-250877>

<http://money.cnn.com/2014/10/10/technology/google-forgotten/>

<sup>91</sup> <http://www.newyorker.com/magazine/2014/09/29/solace-oblivion>

<sup>92</sup> [http://www.huffingtonpost.com/john-p-david/should-the-us-have-google\\_b\\_5420490.html](http://www.huffingtonpost.com/john-p-david/should-the-us-have-google_b_5420490.html)

法修正第 1 条(言論の自由)がプライバシー保護より勝ること、米国には連邦や州によって各自の法律が存在していること、IT 企業によるロビー活動が盛んであることが<sup>93</sup>、米国における「忘れられる権利」の法的整備が進まない要因となっていると見られている<sup>94</sup>。

## 6 終わりに

ビッグデータは、IT を活用した様々な分野でのビジネスの発展に大きな可能性を与えるものと期待されており、米国でも既に多様な形態のビジネスが生まれてきている。米国連邦政府も、ビッグデータは新しい成長産業を創出するものと期待して支援を行っており、今後も米国では官民あげての取り組みが続くものと思われる。

他方、ビッグデータを活用したビジネスが大きくなればなるほど顕在化してくるのが、プライバシーの問題である。新しいビジネスが生まれる時の規制のあり方は、その後の産業発展において重要な点であり、厳しすぎる規制は成長産業の芽を摘み、他方、規制が少ないために事故が多発することも、産業の成長を阻害するという意味では、成長を促す適正な規制が求められる。

我が国でも、これからビッグデータを活用した新しいビジネスが次々と創出されるものと思われる。これらを成長産業に育て、そして我々の生活やビジネスのスタイルを大きく変革させるイノベーションにつなげていくことが期待される。その際、米国での状況を見つつも、我が国の状況に即した取り組みを進めていくことが大切ではないだろうか。

※ 本レポートは、注記した参考資料等を利用して作成しているものであり、本レポートの内容に関しては、その有用性、正確性、知的財産権の不侵害等の一切について、執筆者及び執筆者が所属する組織が如何なる保証をするものでもありません。また、本レポートの読者が、本レポート内の情報の利用によって損害を被った場合も、執筆者及び執筆者が所属する組織が如何なる責任を負うものでもありません。

<sup>93</sup> Google 社は 2013 年のロビー活動に 1,580 万ドルを費やしている。

<sup>94</sup> <http://time.com/98554/right-to-be-forgotten/>