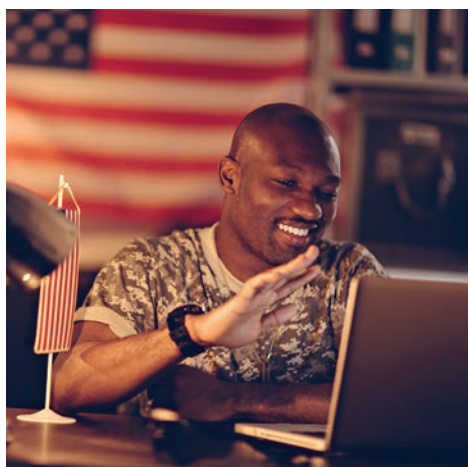
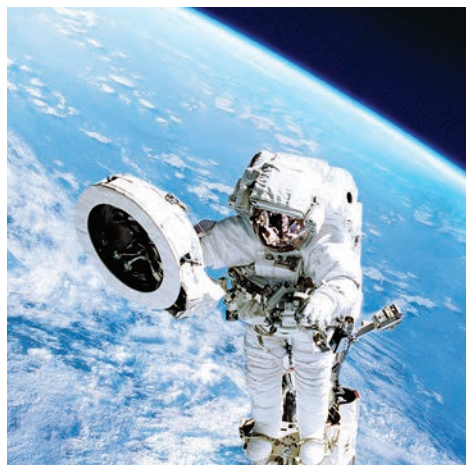


The
Software
Alliance

BSA

データは何をもたらすのか？

～データイノベーションが実現する世界～



エグゼクティブサマリー

ソフトウェアの革新は前例のない進歩を続け、私たちを取り巻く世界を変え、人々に活力を与え、そして、経済を成長させています。

しかし、このデジタル化による可能性をどれだけ引き出せるかは、技術革新により解き放たれるデータが持つ力をいかに活用することができるかにかかっています。私たちは、まさにデータイノベーションの中を生活しているのです。この動きは、データ量の増加だけではなく、これまでの情報収集、蓄積、分析、変換の方法を変えていく基盤技術の進化により引き起こされています。

今日、9割のビジネスリーダーが、データは、キーとなる経営資源の一つであり、土地、労働力、資金と並ぶ事業差別化の基礎的要因であると捉えています。

例えば、つい最近までは、降雨パターンの識別に必要なデータ収集は、数百年に渡る天候パターンの観察を要していました。また、交通網計画には道路脇に座り、走行速度を記録することが求められていました。同様に、病気の原因と治療法を研究するためには途方もない数の手書きノートが必要でした。

今、何百万もの機器、機械、車両、そして街灯に取り付けられたセンサーがあらゆる時に、あらゆる場所でデータを収集しています。かつては莫大な費用がかかるため保持が難しかった大量なデータも、ストレージ容量の増大とコストの大幅な低下により、蓄積されたデータは再生可能資源へと変貌しました。データが多様な目的のために再利用される能力を得たことで、私たちは繰り返し新しい方法でデータを分析、変換することができるようになり、時間、費用の節約だけでなく、生命までも救う貴重な知見を得ることが可能となりました。

収集されたデータには、個人情報が含まれることもあり、最先端のセキュリティと責任ある管理モデルにより、これらに対する安全かつ公正な使用方法が確保されなければなりません。しかし、データの大多数は、機器同士のやりとりの記録、または機器や機械の実行に要されるものです。製造工場の組立ラインから飛行中のジェット旅客機に渡る多くのモノにより、何百万バイトものデータが生成、分析されています。これにより、かつては想像もできなかった手法により生産性を向上させることが可能になりました。

データがあらゆる場所でユビキタスに存在し、使用され、私たちの生活を改善する一方で、多くの人はデータが何か、どこから来て、どの様に使用され、そしてデータが本質的に持つ大きな可能性について理解していません。

本調査報告書は、世界中の非常に困難な問題解決の現場において、データイノベーションがもたらす革新的な進歩につき、いくつかの具体的事例とともに解説します。また、データの収集、分析、変換方法における抜本的な変化により、21世紀そしてその先のデジタル経済において可能となる多くのことが始まりつつあることを示します。そして、データイノベーションを起こす産業が拡大するにつれ広まった誤解について説明し、真実をお伝えします。最後に、本報告書は、データ経済を初めて理解しようとする方々向けに、データイノベーションで使われる用語を定義する用語集を掲載しています。

データイノベーションは、これまでに例を見ない機会を世界にもたらします。革新的なソフトウェアツールが、大量に増加し続けるデータ資源に隠された問題の答えを見つけ出す助けとなっています。これらの斬新なツールが、データから、私たちの生活を変える新しい製品、ソリューションそしてイノベーションを創り出しています。経済的には、データをより上手に活用することで、この先4年間だけで1.6兆ドルに上るデータ配当を導き出すことができます。経済学者達は、データにより効率性が向上することで得られた利益が、2030年までに世界GDPにおよそ15兆ドルも追加貢献すると予測しています。

今日、私たちが賢明な選択をすることで、来たる「データ集中型」経済が、新しい雇用、産業、打開策、治療法等の強力な発生装置として、今後数十年の経済成長を促すことになるでしょう。

『データイノベーション』の定義

データイノベーションの真髄は、「四つのV」にあります。四つのVとは、データの量 (Volume)」、データが生成される速度 (Velocity)」、データの種類 (Variety)」、そしてデータの正確性 (Veracity)です。しかし、生データそのものでは価値がほとんどないこと、そこから真の価値を引き出し、最大化するための大変革を起こすような機会が何であるかについては余り議論がされてきませんでした。

本調査結果で明らかになったのは、日常的なことから生命に関わることまで様々な利益をもたらすために、データは収集、蓄積、分析そして変換されなければならないということです。これらの処理が、データイノベーション、すなわち、膨大な非生産的情報から非常に有益なデータを引き出すことの中核部分です。

目次

5	はじめに
7	データを答えに変換する
14	データ駆動型経済
17	データに関する誤解を真実と切り離す
28	デジタル用語集：データの用語を理解する
34	後注
40	BSA について

紀元前 6000 年

15 世紀

1850 年代

21 世紀

データマイルストーン

はじめに

人類史を通して、文明発展の歴史は、私たちの情報を観察、収集する能力の進歩と密接に関連していると言えます。私たちの祖先は、距離、重さ、量、温度、時間や場所などを測るツールを開発しました。時を経てこれらの技術が改善され、また私たちが狩猟生活から農耕生活、そして都市住民へと変遷する過程で非常に重要な意味を持ちました。

紀元前 6,000 年頃、人類はすでに、農作物の収穫量を増やし、より多くの人に食料を供給するために、収穫量や農作の休閑に関する情報を用いていました。15 世紀には、空からのデータを活用して世界を航行し、公海を開き世界貿易の場を広げました。1850 年代には、データを用いてコレラ蔓延の理由を悪質な水に関連づけて人々の生命を救いました。

近代に入ると、まだ量が限られていたにも関わらず、データは、人類が直面した困難な問題に対し、思いもしない解決へと導くヒントを私たちに与えてくれました。石碑やパピルス、絵巻、印刷物等、何に記録が残されていたかに関わらず、時を経て、広く普及し、注目を増してきたデータは、経済そして人類の進化を支える推進力の一つとしての役割を担い続けています。

21 世紀の今日、この一連の変化が急速に加速しています。データ量が増加し、データストレージ費用が低廉化する中、最新鋭のツールを活用したデータサイエンティストが膨大な量のデータから役立つ知見を明らかにする時代となりました。データ処理技術が更に進歩することで、彼らのもたらす影響はより大きくなり、機会が拡大しています。

私たちは今、境界無き情報と無限の可能性に溢れる世界に向かっていきます。私たちの生活向上に関わる予測をするためにデータはどのように使われているのでしょうか。予測データは、職場への通勤に傘の携帯が必要か、バスに乗る必要があるかなどを事前に知らせてくれます。交通量データは信号機と連動し、電車の到着時間を予想し、例えば子どもの行事に間に合う最短ルートを提示する目的で使用されています。ウェアラブルデバイスと称される身に付けて使う機器は、私たち個人のフィットネス関連の情報を追跡し、長生きに繋がるより賢明な選択をしたり、健康的な生活を促してくれます。そして、科学者はテラ

バイトに上る多大な遺伝情報を分析し、新しい治療や、さらに効率的で一人一人にあった療法の開発を行っています。

データの力

- + **バルセロナ**は、データを活用したスマートシティを構築しています。観光客の往来パターンを調査することで、公共駐輪場の増設場所を検討したり、街のどこに ATM を設置すべきかを特定するのに活用しています。
- + **アラブ首長国連邦**では、最新のデータツールを使用して、世界初のポジティブエネルギービルディングを設計中です。消費するより多くのエネルギーを生産する建物です。
- + **ケニア**では、モバイルデータを使用することでマラリアの感染パターンを解析し、政府による撲滅活動を最適化するための場所特定に役立っています。
- + **アイオワ州からインド**に至るまでの農家は、種苗、人工衛星、センサーやトラクターの情報を活用し、作物の種類、植え付け時期、農場から消費者に届くまでの鮮度の追跡方法、そして気候変化への対応について、より効果的な意思決定を行っています。

収集

蓄積

分析

変換及び転換

自動車を購入する際には、その販売価格を見るだけでなく、データにより、車の燃費、メンテナンス、保険、安全履歴といった有用な情報が得られ、その情報に基づいた判断が可能になります。また、今日の自動車はスーパーコンピューターに車輪を付けたようなものであると言えます。搭載されたプロセッサはパフォーマンスを分析するセンサーと交信することで、オイル交換時期をドライバーに知らせ、電気モーターへの切り替えタイミングを判断し、またバックする際に子供が後ろにいないかどうか知らせてくれたりします。

既に、増加し続ける膨大なデータは、私たちに指先の操作だけで得られる大きな力を与えています。

しかし、そもそも「データ」とは何なのでしょう。誰が、そして何が作り出しているのでしょうか。私たちの生活を改善する可能性とは、一体何のことを指しているのでしょうか。最大限の利益を得るためには、どのように使用される必要があるのでしょうか。そして、どうしたら、私たちの価値観や憂慮を反映した方法で使用されていると確認できるのでしょうか。

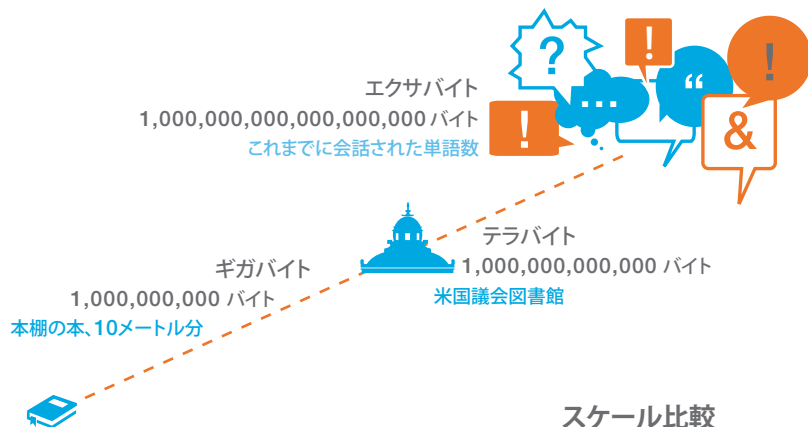
かつては少量であったデータが、膨大で有益な再生可能資源へと変遷し、今や経済、社会的利益の主要資源となった今、これらは重要な質問であるといえます。歴史的にみて、土地や労働力、資金などの資源へのアクセスが成功者と失敗者を隔てる経営差別化要素でした。今日、9割のビジネスリーダーが、データはキーとなる経営資源の一つであり、土地や労働力、資金と並ぶ事業差別化の基礎的要因であると捉えています。¹

一つの例は、効率的なデータ活用が産業における効率性を1%でも改善することに成功すれば、2030年までに全世界のGDPに少なくとも約15兆ドルもの経済効果をもたらすことができると経済学者が見積もっていることです。²「次の大きな出来事」は、インターネットと接続された、私たちを取り巻く世界に関するより正確なデータを生成する何10億もの小さなモノが、より強力なデータ駆動型ソリューションにより実現するかもしれません。³私たちは既に、抱いていたことすら知らなかった疑問への答えを見つけ出しているのです。

この大きな変化は、既に進行中です。人々のあらゆる活動がデータを生成し、そして全く新しいデータの流れが毎日創りだされています。世界に存在するデータの90%以上は過去2年以内に作成されたもので、データの生成速度は二年毎に倍増しています。生成されるデータのほとんどは、個人情報に該当しません。この点を理解することは重要です。プライバシーを保護することが不可欠である一方で、生活改善に役立つデータの多くは機械に取り付けられたセンサーが生成しているのです。

私たちの課題は、データに秘められた貴重な知識を理解するための創意工夫をして、データを活用することです。それは、データを処理し、観察されることを知識に変換し、知識を解決策に変換し、今日の重要な課題に対する有益なソリューションを生み出す能力を指しています。

企業と政府が一丸となり
イノベーションの原動力を
駆動させる必要があります。



データを解決策に

変換する四つの ステップ

今日のデータ革命を煽っているのは、これまでのデータ収集・蓄積・分析・変換の方法を変えていく基盤技術です。これらの技術が互いに相まって、データの奥深くから有益な知見を探しだし、新たな知識を解き放ち、新たな関係性を見出し、新たな予測を実現します。

1

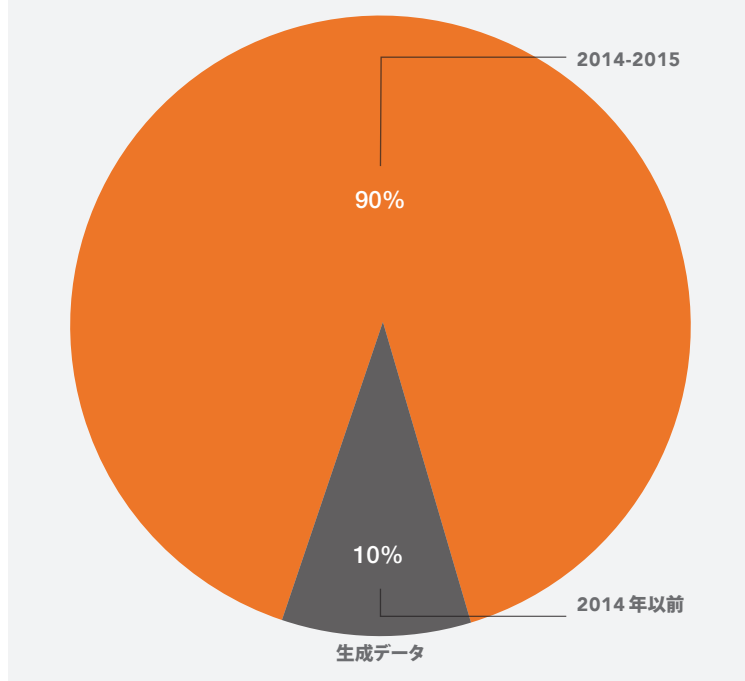
データの収集

有史以来、私たちは常にデータを収集し、社会の発展に役立ててきました。しかし、これまではデータ量が非常に限定的でした。今日、私たちは、身の回りから、より有益で意味あるデータを使いやすい形式で収集し、意図した結果をもたらす重要なデータを活用できるようになりました。データは、かつての限定的な資源から、非常に豊富で、重要かつ再生可能な資源へと変化を遂げたのです。

これは、個人情報に関連付け、抽出することによって可能になったのではなく、様々な機器、センサーがインターネットに接続され、それが指数関数的な勢いで成長する新しいデータを大量生成することで実現しているのです。その結果、交通量測定用の路側センサーや、デジタル音楽、映画、地球を周回する人工衛星、工場や金融市場を制御するセンサーやシステム、次世代の新しいモノをデジタルな手法で創り出すツールなど、様々な方法によりデータはあらゆる場所で生成されるようになりました。情報は、より速く成長し、遠くに広がり、そして重要性を増しています。

今日の世界におけるデータ増加量

世界におけるデータの90%が最近2年以内に生成されたものです。



情報元:
IBM <http://www.ibm.com/software/data/bigdata/what-is-big-data.html>

毎日、250京バイトに上るデータが生成されていると推定されます。⁴アナログ時代の感覚では、このデータの莫大さは想像することすらできません。分かりやすい例えを使うと、昨年一年間だけで世界全体が生成したデジタルデータを記録したDVDを重ねると地球から月を往復してしまいます。⁵データ生成の速度はさらに加速しています。世界中の企業が扱うビジネスデータ量は、1.2年毎に二倍に膨れ上がっています。⁶これらのデータはいったいどこから発生しているのでしょうか。以下は、数多くあるソースの例の一部です。

データの多くは、個人を特定できるものではありません。近い将来、地球全土に接続されたデバイスが私たち自身を取り巻く世界をさらに深く理解し、改善する助けとなってくれるでしょう。

- + 診断画像など、病院におけるデジタル情報の増加量は2015年までに毎日665テラバイトに達すると見込まれ、これらは治療方法の発見や救命に役だっています。⁷
- + 大陸間を飛行する今日の旅客機は、エンジンやフラップ、着陸装置といった部分に接続されたセンサーであふれており、これらは毎フライト1/2テラバイトものデータを生成し、フライト性能を改善し、⁸乱気流を避け、安全性を向上し、そして従来より2,000倍以上速くエンジン不良可能性を特定することを可能としています。⁹ 毎日のフライト数が25,000便以上あることを踏まえると、旅客機だけでも非常に有用なデータが多く生成されているということがわかります。
- + 気象衛星、天文气象台、レーダーその他のセンサーが22.5億を超える気象データポイントを毎時15回、つまり毎日20テラバイトものデータを収集し、世界中の正確な気象予測を可能にしています。¹⁰
- + 金融取引所は毎日4-5テラバイトに上るデータを生成し、リアルタイム分析、取引活動の監視、ビジネス成長、そして豊かな経済成長を助けています。¹¹
- + 何万もの配送車両に搭載されたテレマティックセンサーは、エンジン性能、ルート改善、問題の事前予測を行っています。マッピングデータ分析と組み合わせた車両センサーデータにより、企業は何百万ガロンの燃料を削減し、年間何千台もの車両の路上排出量に相当する排気ガスを削減することに成功しました。¹²
- + 欧州原子核研究機構（CERN）の大型ハドロン衝突型加速器は、宇宙がどのように動いているのか、その大いなる秘密に迫る新しい見識を見出すため、実験ごとに毎秒40テラバイトものデータを生成しています。¹³ 同様に、チリの大型シノプティック・サーベイ望遠鏡は毎晩の観測で、宇宙に関する30テラバイトものデータを生成します。¹⁴

- + 単一のDNAゲノムの配列決定だけでも、200ギガバイトのデータを生成します。DNA配列のコストが下がること

で、科学者は、医療におけるブレイクスルーにつながる相違点、類似点を見つけて、生命を救うために何十万もの配列が格納された莫大なデータベースを構築しています。¹⁵

指数関数的に増加しているデータ量だけでなく、それを生成する方法も同様に増えています。私たちの周りの環境とインターネットを接続する機器の数が増え、「モノのインターネット」を作り出すことで、数多くのセンサーが毎日、全く新しい形のデータを創りだしています。非常に強力なセンサーを備えた推定500億ものデバイスが2020年までにインターネットに接続されることが計画されており、「次の大きな出来事」は非常に小さな物事に基づく発見となるかもしれません。¹⁶

これらのデバイスは、土壌水分、エンジン性能、エネルギーシステムの効率性、および喘息発作の場所などの測定を通じてデータを作成します。人間である私たちは、五感を使用して自身を取り巻く世界を理解します。近い将来、地球全土に接続されたデバイスが、物理的世界における特徴すべてを理解し、私たち自身を取り巻く世界をさらに深く理解し、改善する助けとなってくれるでしょう。そして、私たちはさらに新しく有益で膨大なデータを生成するのです。

データストレージのコストが下がり続け、データ量が成長し続けているため、データの用途は拡大し続けています。

2

データの蓄積

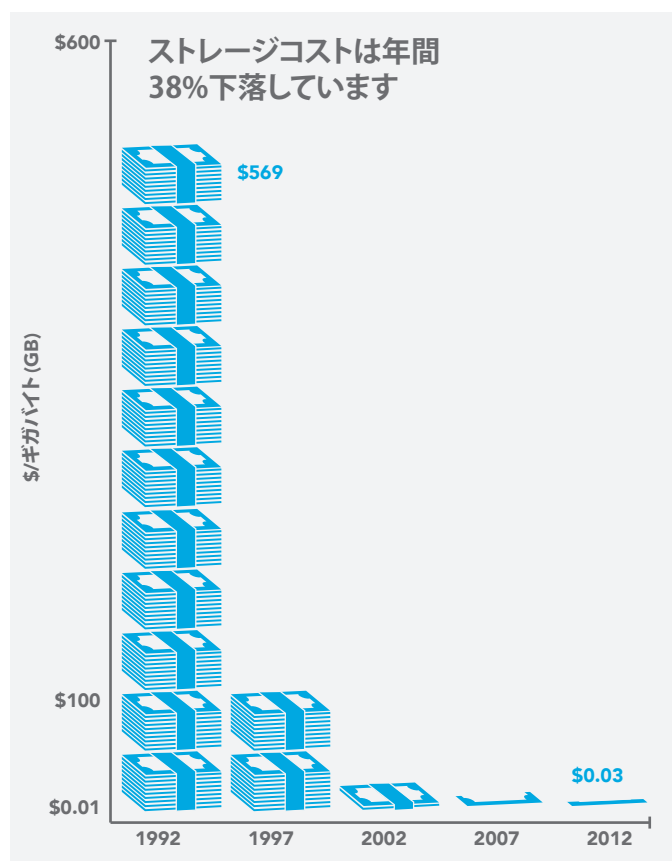
ストレージのコスト下落が、データ駆動型イノベーションを実現しています。1980年には、ギガバイト相当のデータストレージでさえ入手困難とされ、何百何千ドルという費用がかかり、管理するためにフルタイムで働く人材を必要としました。¹⁷ 今日、ギガバイトのストレージは、たった数百円のコストで手に入る上、管理は簡単、そして、いつでもどこでもアクセスできるようになりました。¹⁸ 1980年代以降、ストレージの価格は千万倍も下落しました。¹⁹ 分かりやすい例えで説明すると、ガソリン価格が同じ額だけ下落したとすると、1980年に支払ったガソリンの量で世界を一万回も走りまわることができるのです。²⁰

ストレージのコストは下がり続けているので、さらに多くのデータ量を蓄積することができます。1994年には、世界のデータのわずか3%のみデジタル形式で蓄積可能でした。²¹ しかし、2007年までには94%ものデータがデジタル形式で蓄積されました。²²

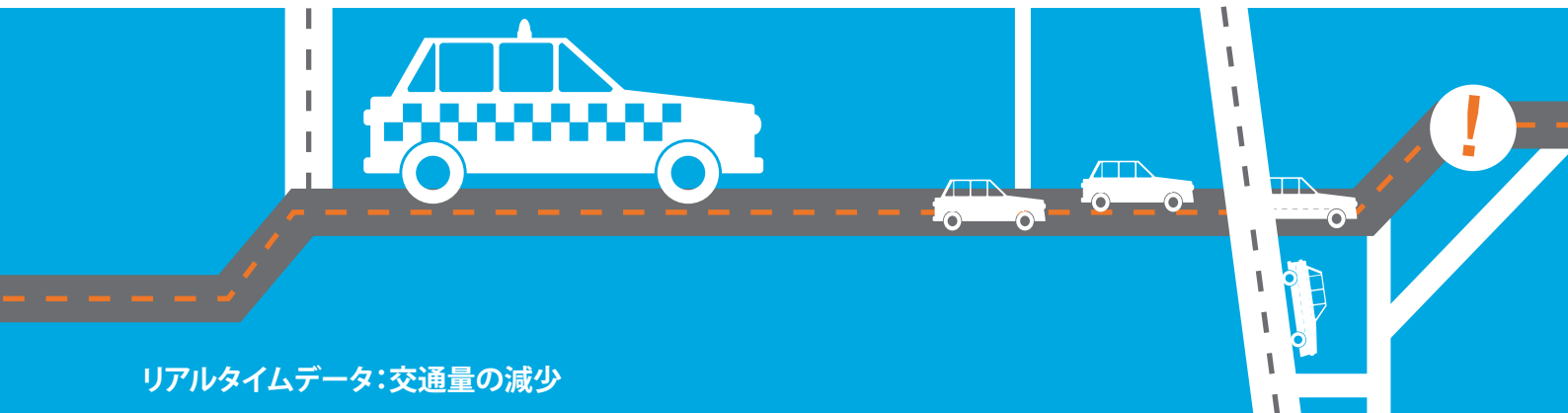
データストレージのコストが下がり続けデータ量が成長し続けているため、データの用途は拡大し続けています。データを遠隔に蓄積し、様々なデバイスからアクセスできるクラウド技術が、データストレージの価格を劇的に下げ、それにより新規データは一度蓄積されると削除されることがなくなりました。²³

つまり、そのほかのリソースと異なり、データは一度使われたら終わりの使い切りではないのです。データは、他のデータセットと結合できる再生可能資源で、データが最初に作成されたときに予期せぬ疑問が発生しても解決策を導き出すために何度も使用することができます。例えば、気象データは、私たちが傘を持つ必要があるかどうかを予測するためにのみ使用されるのではなく、作物の収穫量予測にも活用されています。

クラウド技術により、ユーザーは品質、信頼性、柔軟性の高いデータを手頃な価格で利用することが可能になり、膨大な量のデータを社内ITインフラストラクチャに保管する面倒から解放されました。データストレージの売買、配送方法が根本的に変わったことで、またいつでもどこからでも仮想的にデータが利用可能となったことで、ここ10年内で最も変革を促す技術の一つとして、またデータ駆動型ソリューションを実現する素晴らしい技術の一つとしてクラウド技術が発展しているのです。



情報元: Hagel III, John et al. From Exponential Technologies to Exponential Innovation. Deloitte University Press, 2013. 印刷物: 2013 Shift Index Series.



リアルタイムデータ：交通量の減少

3

データの分析

データは、理解されてこそ価値があり、そうでなければ観測データの単なる寄せ集めです。データに含まれる知識を理解することは、人間の創意工夫と革新的ソフトウェアを組み合わせることによりのみ実現可能です。

あらゆることがますます自律式になる世の中にもかかわらず、データに潜む解決策を解き放つために要されるのは個人の興味や、人間のスキル、集中的な作業です。

生データを活用するには、先ず整理する必要があります。一説では、データサイエンティストはたった少しの有益なデータのために混沌とした状態のデジタルデータに50%から80%ほどの時間を割いていると言われています。²⁴

そして、適切な疑問を投げかけ、不良データを分類、認識し意味のある方法で結果を解釈することで解決策を導きだせるかは、人の創造性によります。データサイエンティストの役割は、アナリストであり、アーティストであり、そしてストーリーテラーでもあると描写されています。²⁵ 個々のデータ一片は、画面上のピクセルのようなものです。単独では、本当にわずかな情報しか得られません。しかし、正しく並べられた大きなピクセルデータと組み合わせた時には、データサイエンティストは千の言葉程意味のある絵を描き、新しくかつ予想外の意義をデータから導き出すことが可能です。

データをふるいにかけることで、分析ツールは、データの山を切り開き、ユーザーに新しいパターンや傾向を発見する手助けをするばかりか、見かけには関連の無さそうなデータから予想外の見識を発見したり、統計的に興味深い関連性を自動的に発見するのに役立ちます。ますます豊富になるデータベースと絶えず進歩する統計的アルゴリズムを使用するソフトウェア分析ツールにより、私たちは山の様なデータを取捨選択し貴重な情報の宝を見つけることができます。

幸いなことに、創意あふれるソフトウェアと組み合わせることで今日のコンピューターはより強力な処理能力を実現しており、最新鋭のツールを備えたデータサイエンティ

ストは膨大な量のデータを理解し、その中に潜む貴重な見識を発見することができるようになりました。

今日のネットワークは目を見張るものがありますが、一方で、一括ですべての処理を行うためにネットワーク全体の膨大な量のデータを一ヶ所に移すことは、しばしば、経済的に追いつかないことに加え、ロジスティック的側面からも不可能であると言えます。しかし、今日利用可能なより強力な分析エンジンのいくつかは、大規模な並列型分散クラウド・コンピューティング・プラットフォームによって手頃な価格で実現可能になってきています。これらのプラットフォームにより、ユーザーは複数の場所に蓄積されているデータ全体に対し一度に世界的規模のデータ分析ツールを実行することができます。

このデータ分析によって私たちには何が可能になるのでしょうか。かつては信じ難かった未来予測が、今や必然に思えます。今日、最新のデータ分析技術の恩恵により、いつでも信頼性の高い予測を立てることが可能になりました。天気予報も、10日先程度の予測であれば、信頼性が大きく向上しています。車両管理者は、車が故障する前に、どのエンジンが修理を必要としているか予測することができます。現在と過去のデータを比較できれば、多くの場合、未来予測に使用できます。

経済学者は、市場や雇用、インフレをより精度高く予測する方法を探しています。長い間、政府発行の経済統計データが、意思決定者に過去ばかりを見させてきました。GDP成長率といった政府の経済統計は、数か月前の情報を提供しており、大幅に遅れて入手可能となる経済の過去の動きを、将来を予見する最良のベンチマークであるとして伝えているだけです。今、経済学者は、新しい求人情報や業界の受注状況といった様々なリアルタイムデータを組み合わせ、過去のデータと比較し、日々の動態が理解できるより正確なイメージを描き、そして健全な経済を実現するためのより良い政策を策定しています。



IBM とストックホルム市が提携し、1600 個の GPS をタクシーに取り付けました。GPS 機器からのデータは IBM ストリーミングソフトウェアを使用して高速処理され、交通量や移動時間、最適な交通ルートの算出に利用されています。

排気ガスの
減少量：

10%

都市の交通量減少度：

20%

平均移動所要
時間の減少度：

50%

非課税の地球
に優しい車両の
割合の増加：

9%

情報元：Bertolucci, Jeff. 'Dublin Points Big Data Tech At Traffic Jams'. InformationWeek 2013. ウェブサイト。
Nusca, Andrew. 'Stockholm Uses Real-Time GPS Data To Manage Traffic Congestion ZDNet, 2010. ウェブサイト。

リアルタイムデータ分析の向上は自律的な意思決定を実現し、私たちそして私たちが動かす機械による、以前より迅速かつ緻密な意思決定を可能としています。すでに、アメリカの大手自動車会社は、数百のセンサー、テレマティクス、リアルタイム接続機能を備えた新しい車を設計しており、自律駐車などの進歩を実現しています。自動車メーカーはまた、自律式衝突回避や自動運転を実現するリアルタイム分析ツールを推進しています。これらの進歩は、いつか、人間より速く、より確実に状況対応をすることで生命を救うかもしれません。

より強力な処理能力を備える今日のコンピューターを革新的なソフトウェアと組み合わせることで、最先端ツールを使用するデータサイエンティストは膨大な量のデータを理解し、その中に潜む貴重な見識を発見することができるようになりました。

指数関数的に成長する、身の周りの環境を扱うリアルタイムデータにより、学習とほぼ同時にその内容を理解できれば、データ分析ツールの効果を最大限に引き出すことができます。今日、最高のツールは、データの奥深くに潜む新しい相関性を作り、人が正しい尋ね方に気づかなくても予想外の解決策を見つける能力にあります。世界中で、分析ツールは、影響力のある相関関係を見つけ出し、予想外の結果をもたらしています。例：

- + カナダ人研究者は、毎秒1,000以上のデータポイントを追跡し、関連づけることで、異常に安定した生命徴候を示す未熟児は、その翌日に深刻な発熱を発することを示し、医師たちに衝撃を与えました。これにより、医師が予防処置を取ることができました。²⁶
- + 過去の二十年に渡り蓄えられた新聞の情報が、アンゴラ等の地域で、いつどこでコレラが流行するかを予測するために使用されています。²⁷
- + 警察当局は当初地震を予測するために設計されたアルゴリズムを改変し、今では犯罪が起こりそうなエリアを半径150mの精度で予測するために活用しています。このソフトウェアが使用されている地域では、強盗が33%減少し、暴力犯罪は21%も減少しています。²⁸
- + データアナリティクスおよび波、潮流、その他のデータをモニタする海洋センサーの活用により、研究者はデータ分析を使用して、津波やその他の自然災害、そしてその影響に関する妥当な予測をしています。²⁹
- + 医師による訪問回診と処方箋情報から得たデータにより、自己免疫疾患を患う患者ががん発症の高いリスクにさらされていることが明らかになりました。³⁰
- + クレジットスコアデータは、どの患者が処方された薬を摂取するために「親切的な」リマインダーを必要としているかを予測するために使用されています。³¹
- + 気象パターンと相関付けられた十年間分のフライト履歴データを活用して、航空旅客者はどの便が時間通りに到着するか予測することができます。³²

データが私たちを取り巻いているということは、同様にイノベーションの機会も私たちの周りを取り巻いていると言えます。イノベーションが独創的かつ責任ある行動をとることで、データイノベーションが日常の問題から世界が窮する大きな課題まで様々な問題の解決策を与えます。

4

データの変換と転換

強力な最新ソフトウェアツールを使用して、私たちは、データを活用し、感覚や直感ではなく事実に基づいたより良い意思決定をする能力を得ることができます。

具体的には、新しいツールセットは、データを推定、集中、視覚化、反映、洗練、モデル化、予測のために変換し、活用目的に合わせて加工します。こうしたツールには、データを理解する機械学習技術、シナリオをテストしデータを現実世界のソリューションに適用させるモデリング・シミュレーション技術、そして音、画像、映像を認識し、より意味のある形式に変換するツール等が含まれます。

このようにデータを変換することは、計画策定、設計、意思決定プロセスの改善につながります。例えば、今日、医師は大量な最新の研究成果に圧倒されており、リアルタイムで患者のデータを捕捉することも合わせ、処理しきれない状態に置かれています。³³そこで、病院は、臨床意思決定支援システムに目を向けています。複雑なデータ環境において、より迅速に信頼性のある診断を下すため、様々なソースから得られるデータを分析するソフトウェアシステムを使用することで、70%以上の事例で有益であることが証明されています。³⁴

他にもソフトウェアツールが、データを転換し、より意味のある形にする事例があります。音声、画像、映像データのリアルタイム処理は、人々の生活を大きく変えています。例えば、人の会話に関するデータがより多く収集されたことで、音声認識技術が継続的に改善されてきました。これにより、外国との音声会話のリアルタイム翻訳は飛躍的進歩を遂げ、今後、グローバルな商取引、貿易の機会を広げる可能性があります。同様に、難聴患者が約3.6億人い

る中で、中国の研究者は、パターン認知と3D Kinect センサーから得られるデータのリアルタイム処理を活用し、手話と話し言葉、書き言葉の間を双方向にリアルタイム通訳するシステムを開発しています。³⁵

画像処理の高速化は、癌の検出、認知コンピューティング、神経生物学、ロボット工学などの分野で大きな影響を与えています。例えば、予測不可能な外観や形状が特徴の脳腫瘍を、医療現場の画像診断で特定するのは特に難しいと言われています。クラウド・コンピューティングおよび高度な画像解析アルゴリズムの力を借りて、科学者のチームは競って、より正確かつ迅速に脳腫瘍を識別するための最良のソフトウェアアルゴリズムの発見に取り組んでいます。³⁶

マンモグラフィ用画像が二次元から三次元に移行することで、乳癌の検出率が向上しています。三次元的マンモグラフィは、ソフトウェアを活用して異なる角度のX線画像を組み合わせることで三次元画像を生成し、乳癌の検出率を向上させ、誤診を低減します。³⁷

視覚化とシミュレーションを活用することにより、データの理解と使用が容易になります。複雑なシステムをモデル化してシミュレーションをすることで、実際に構築することなく、高い精度と速度を備えたテスト設計が可能になりました。例えば、1980年代に、ボーイング社ではボーイング767の開発において、77機ものプロトタイプを物理的な風洞を使用してテストする必要がありました。2005年までに、同社は時間、エネルギー、費用削減及び安全性向上のために仮想風洞とスーパーコンピューティングを活用したことで、ボーイング787の開発においては、わずか11回の物理テストしか要しませんでした。³⁸



仮想風洞は、流体力学における3D計算を簡単かつ迅速に使用するために用いられる大量データ高速処理ツールの一例に過ぎません。これらのツールは、熱流量、流体の流れ、空気の流れ、プロセスフローについて優れたモデル化を提供することでパフォーマンス向上に活用されています。モデル化されたデータは、汚染物質が地下水を通じて拡散状況の把握、

風力タービンの性能向上、自然災害に耐える建築物の設計などに使用されます。

要するに、これらのツールは、データをソリューションに変換するのです。

データを収集、蓄積、分析、変換する機能が組み合わせることで、さらに素晴らしいソリューションを生み出す新しい機会が見えてきます。

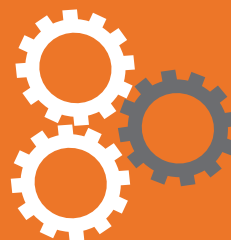
しかし実際には、これら四つの機能は多くの場合、別々の人により実行され、多様なデータセットに対して処理され、分散した場所に蓄積されています。これはデータ革命が持つ力の一部に過ぎません。関連性を有しなかった別々のデータが、異なる場所に蓄積されていても統合、分析することができ、データが構造化されていなくても解釈され、データ作成者が知る由もなかった疑問に対する抜本的な解決策を明らかにするために活用されるのです。データセットに含まれるデータが作成された時には、そこから抽出できる情報にどんな力が秘められているか思いもよらないことがよくあります。見かけは全く関係なさそうなデータの組み合わせが、後になって価値を見出したり、知り得た知識を活用するためのアルゴリズムが未だ発明されていなかったりするからです。

データが私たちを取り巻いているということは、同様にイノベーションの機会も私たちの周りを取り巻いていると言えます。イノベーターが独創的かつ責任ある行動を取ることで、データイノベーションが日常の問題から世界が窮する大きな課題まで様々な問題の解決策を与えます。生産性、経済成長そして個人の利益の新たな波を加速させ、データが実現する可能性を最大限にするために、私たちは再びイノベーションエンジンを駆動させ、最先端のソリューションとそれがもたらす成果を受け入れる準備をする必要があります。



データを最大限活用する製造会社の、
4年間に渡る

**コスト削減額
は3710億ドル**



ビジネスデータ:コスト削減

データ駆動型経済

データは現在、経済的利益をもたらす最も強力な要素として浮上しています。データが収益にもたらす影響は、ハイテク及びローテク業種双方のあらゆる分野を通じて波紋を広げています。今後4年間で、データを効果的に活用することにより1.6兆ドルに上る「データ配当」が得られると見積もられています。³⁹

データイノベーションは生産性を向上する能力を実証してきました。データ指向型の意思決定を既に活用している企業は、生産性が5%から6%向上したことを報告しています。⁴⁰ データを効果的に活用することで、産業全体の効率性が1%でも改善すれば、2030年までに全世界のGDPに約15兆ドルもの経済効果をもたらすと経済学者は予測しています。これは、今のアメリカ経済と同規模の経済が追加されることを意味します。1%の生産性の向上は小さいように見えますが、ゼネラル・エレクトリック社CEOジェフ・イメルトは、「石油を扱う人にソフトウェアで何かを1%節約してあげようと伝えたら、彼は一生涯の友になることでしょう」と表現しています。⁴¹

データイノベーションは雇用を創出する

データイノベーションは経済成長を後押しするだけにとどまりません。新しい雇用機会を創出する強力な原動力となり、雇用を倍増させる有力な要素となります。データイノベーションは、データアナリスト、ソフトウェア開発者から、データイノベーションを実現する巨大なデータウェアハウスを運営する者まで、既に何千もの高賃金の雇用を創出しています。大手IT調査企業ガートナーの調べによると、今後4年間に440万の雇用がデータイノベーションを支えるために創出されると見込んでいます。またデータ関連のIT系雇用1人あたり、IT外雇用3人が創出され、経済全体で数百万人以上の新たな雇用が生まれると考えられています。⁴² さらに、米国で61%、欧州で58%の上級役員がデータアナリティクスが自社の雇用計画にとって重要であり、新たに採用を行うと言っています。⁴³

米国では

データ関連

の雇用創出1件あたり新たな

間接雇用3件

を生み出します。

61%

の米国の上級役員が、

データアナリティクス

が自社の

**雇用計画にとって
重要であると述べて
います。**



スマートビルディングは単独で

**年間250
億ドル**

エネルギーコストを
削減します。



今後15年間の世界における
1%のデータ駆動型の生産性
向上が航空業界で

**300
億ドル**

の燃料費用削減を
実現します。



データは、ビジネスの機動性、応答性、競争力をさらに高めます。

かつては大規模なIT駆動型の生産性向上により経済を成長させ、日常生活の水準を引き上げたのは、テクノロジー業界など少数の業界に限られていました。しかし、多くの業界が多様な形式のデータを生成するにつれ、これまで遅れていると言われていた農業からヘルスケア、運輸から教育、エネルギーから金融まで全ての業界がデータによる利益を享受しています。実際、最近の調査によると米国の79%、欧州の80%の上級役員が、データ分析は、顧客のニーズによりマッチしたサービスを提供するための自社計画に重要であると回答しています。⁴⁴ また、米国では70%、欧州では72%がデータ分析は、新製品やサービス開発に係る自社計画にとって重要であると回答しています。⁴⁵

ヘルスケア

データイノベーションは、人々がより長く健康な生活をするための知見をもたらします。

ヘルスケア業界は、病院毎に一日当たり数百テラバイトものデータが生成されているなど、潤沢にデータが存在する業界の一つです。⁴⁶ このデータが最大限に活用されれば、より良い治療、処置の処方が可能になるでしょう。今日、データは、医師に、患者の健康状態に関し、これまで以上の知見を与え、より良い意思決定を可能にしています。ソフトウェアによる分析の力を借りることで、医師は、過去の患者全てについて情報を活用して、将来の患者全てに対する治療の選択肢を提示することができます。しかし、現状ではヘルスケアに関するデータは十分に活用されているとは言えません。ヘルスケア業界が効率性と品質向上を目的にデータ活用を行えば、費用削減効果は年間3000億ドルに上ると言われ、これは支出全体の8%にあたります。⁴⁷ しかし、最大の効果は、費用削減額ではなく、救われた生命の数で測られるべきでしょう。一例として、研究者は、患者の病歴とのリアルタイムデータを分析することで、66%の精度で心停止を発生 of 4時間前に予測することが可能な機械学習アルゴリズムを開発しました。⁴⁸

交通

データイノベーションは、時間、費用、燃料を節約する他、死亡事故減少に貢献しています。

データにより、未来のスマートな交通の到来が近づいています。航空、鉄道、道路交通を通じ、リアルタイムデータ分析により、物資と人の移動に関する安全性と効率性を向上させるスマート交通システムが実現します。今日の新しい車は、最大毎時25ギガバイトのデータを生成し、データを処理するために1000万行ものソフトウェアコードを含んでいます。⁴⁹ データは、シートベルトに匹敵する安全性を有する衝突回避システムや新しい安全装置の駆動に使用されています。⁵⁰ これにより、怪我と死亡を50%減らすことができたのです。路上の車両数は増加しつづけていますが、データを革新的な方法で使用することにより、路上の渋滞を減らし、あるいは渋滞を回避する経路提示によって、無駄な時間を何百時間も節約し、何千ガロンもの排気ガス、温室効果に繋がる何トンもの排出物の削減に繋がります。

航空データも増え続けています。最新の旅客機に取り付けられたセンサーからはフライト毎に最大0.5テラバイトのデータが生成されます。このデータを使用して飛行性能を向上させ、乱気流の回避および安全性の改善を行い、また従来より2000倍の速さでエンジンの欠陥を認識できます。⁵¹ 航空データは、飛行経路計画を改善し、部品交換の必要性を故障前に知ることができます。これらの利益は、積み重ねられ、航空業界全体でデータ駆動型の生産性が1%向上するだけで、300億ドルの燃料費用削減が可能です。

エネルギー/環境

データイノベーションは、エネルギー消費を減少し、私たちの環境を改善します。

また、多くの業界における、環境改善と、何億ドルものエネルギー節約に貢献します。データから得られる知見が、電力会社、建物、家庭、工場や農場における賢明なエネルギー使用を実現するのです。例えば、データ設計ツールを使用して、インテリジェントシステムと大量のセンサーデータ、分析力、アクチュエータを組み合わせ、効率的な管理を可能にし、エネルギー消費を削減させる建物を設計することが可能です。これらを合わせると、スマートビルディングでは建物だけでエネルギー費用を年間250億円も節約できると言われています。⁵²

ヘルスケア業界が効率性と品質向上を目的にデータ活用を行えば、費用削減効果は年間**3000億ドル**に上ると言われ、これは支出全体の**8%**にあたります。

製造業

データイノベーションは、製品の設計、構築、分散方法を改善します。

製造業は、他のどの業界よりも多くのデータを蓄積します。⁵³そのため、製造メーカーはデータを活用することにより、製品設計、生産、流通の各過程において、生産性、品質を改善する余地が多くあります。一説では、製造業においてデータを効果的に活用することで、製品開発時間と組立コストを50%減少できると言われています。⁵⁴また、IDC（国際データ・コーポレーション）は、データを最大限に活用する製造メーカーは、データから4年間で総額3710億ドルの「配当」を得られると推定しています。⁵⁵リアルタイムデータを用いることで、企業はグローバルサプライチェーンの追跡、管理を改善し、製品の欠陥を減少させることが可能になります。

データは設計改善も支援します。データ駆動型のデジタル設計は、数え切れないほど多数のアイデアをテストし、最終的に、最適な製品デザインに到達するまで、仮想環境で微調整を繰り返すといった設計方法を用いる反復設計プロセスを強化します。トヨタ、フィアット、日産は、このデータとモデリング技術の協働環境を採用し、新モデル開発時間を30%から50%ほど短縮しています。⁵⁶

金融

データイノベーションは、効率性を向上し、コンプライアンスを強化し、不正行為を防止します。

金融サービス部門では、増大する豊富なデータに着目し、業務効率改善、コンプライアンス強化、不正行為特定に活用しています。ある調査では、銀行や金融市場の企業の71%が、情報と分析の使用が、自社の競争優位性を作りだしていると報告されています。⁵⁷データへの投資からは、莫大な「配当」が得られます。あるクレジットカード会社では、年間20億ドル相当のクレジットカード詐欺特定にデータ分析が活用されました。⁵⁸

農業

データイノベーションは限られた資源を活用し、高品質な食料をより多く生産することを可能にします。

データは世界中の農場において、増加する人口に対応する、美味しく栄養価の高い食料を供給するのに役立っています。精密農法という概念に基づき、世界中の農家が、歩留まりを向上させ、コストを削減し、より多くの人々に供給するための優れた意思決定を行うため、種苗、人工衛星、センサー、トラクターから入手したデータを活用しています。世界の人口の4分の1を占める人々が農業と食糧生産に関わっており、データを上手く活用すれば、大きな利益が得られます。例えば、データ分析ツールを活用することにより、農家は種苗にかかる費用や、農薬、化学物質の使用量を削減することができ、1エーカーあたり最大5-10ブッシェルも収穫を増やすことができます。⁵⁹酪農家については、クロアチアのスタートアップ企業がクラウドベースの分析プラットフォームソフトウェアを開発し、飼料品質と生産量の相関関係や受胎率などのリアルタイムデータ活用を可能にすることで、農家の生産性と効率性を50%も向上させました。⁶⁰

カナダ人研究者は、

毎秒

1,000 以上のデータ ポイント

を追跡し、関連づけることで、異常に安定した生命徴候を示す未熟児は、その翌日に深刻な発熱を発することを発見し、これにより医師が予防処置を取ることができ、救命に繋がりました。

カナダ



データにまつわる誤解と真実を 切り離す

近年のデータイノベーションとデータ経済に関して、しばしば誤解が生じることがあります。これらには下記のような誤解が含まれます。

- + 個人情報とデータ保護
- + データ経済がもたらすインパクト
- + データの信頼性
- + 21 世紀のデータイノベーション
- + データイノベーションが世界にもたらす恩恵
- + データの規制における政府の役割

個人情報とデータ保護

誤解

データイノベーションの恩恵を受けるのはIT企業のみであり、個人ではない。

真実

データイノベーションにより、消費者は賢明な選択が可能になり、販売者は顧客に対してより良い商品をより良い形で提供できるよう商品やサービスを工夫できるようになりました。パーソナライズされた教育、パーソナライズされた医療、そしてパーソナライズされたエンターテインメントを可能にする大きな発展をもたらしたのです。また、大量生産型の経済から、マスカスタマイゼーション型の経済への転換が促進されます。確かに、データイノベーションを可能にする革新的な新技術の開発、提供の中心にはいくつかの企業が存在しています。しかし、私たちが賢明な選択をすることにより、データ中心経済が新たな雇用機会や産業を創出し、世界がより消費者側の手で動かせるようになるのです。

誤解

データの全てが個人情報である。

真実

データには、個人情報が含まれる可能性があります（例：モバイル機器の使用により生成されるデータ、ソーシャルネットワークの使用により生成されるデータ等）。しかし、データの多くは個人情報に該当しません。

日々生成される膨大な量のデータには、人工衛星による気象観測情報やジェットエンジンの動作、コンピューターが生成する株式市場取引情報、個人に関係ないセンサー等が生成する情報が含まれます。個人に関係するデータの場合でも、多くの場合は他人によりアクセスされることがなく、あるいは匿名化され、個人の特定につながる個人情報を含まない形で蓄積され、使用されています。

インドでは、インターネット、キオスクが

4百万以上の農民

に作物の価格、天候、そして
その他の情報を

現地語で提供しています。

インド



誤解

企業は個人情報保護について関心がない。

真実

生成された個人データは適切に保護される必要があります。データ活用機会を増やすためには、人々からのデータに対する信頼性と信用度を高く維持する必要があります。データを利用する企業や組織は、データ管理に最善を尽くしてその責任を履行しなければなりません。その実践には業界主導による標準化や、責任あるデータ利用のための自主的ガイドラインの策定が進められています。この分野における多くのリーダーが、既に消費者に関するデータの収集方法と共有方法について自主的に開示しています。また多くの企業が、個人情報匿名化についてのベストプラクティスに倣い、できる限り実践しています。

誤解

データイノベーションにより、プライバシーが全て失われる。

真実

データ経済が成功するか否かは、消費者から得られる信用にかかっています。個々が自分の個人情報が安全であると感じられることが大切です。先駆的なソフトウェア開発者は、開発当初からプライバシー保護をシステムに組み込む「プライバシーバイデザイン」を採用してシステム構築をしています。開発者は、さらに、匿名化、非識別化、暗号化ツール等を活用して、データ漏洩による影響を最小限に抑えるよう努めています。データは、集約されることにより、個人ユーザーを特定できなくても、行動パターン解析が可能であり、ユーザーからの信頼やプライバシーを侵すことなく、有益に活用できます。また、データ漏洩や誤用に対する相対的リスクが高い機微情報（金融やヘルスケアデータ等）に対しては、最高レベルのプライバシー保護を課すよう強制力あるプライバシーポリシーを適用させることができます。これはつまり、気象データやビジネス分析といった個人情報とは関係のないデータは、患者を特定できるヘルスケアデータに課すのと同様の保護レベルを要しないことを意味します。

誤解

データを完全に非識別化することは不可能である。データの非識別化は効果がない。

真実

データの非識別化は、情報が特異的な個人と紐づけられることを防ぐために使用されるプロセスです。データが一度非識別化されれば、個人を特定されることなく分析が可能となります。専門家は、データがプライバシーとデータ品質の両方を最大限活かせる方法で非識別化する技術を開発しました。⁶¹ 専門家によると、非識別化が適切にされると、匿名化されたデータから個人が特定されるリスクは、多くの場合、1%未満であると言われています。⁶²

誤解

データを使用する企業は、信頼できない。

真実

産業界はプライバシーの問題を気にかけ、耳を傾けています。今では、より良いプライバシー保護を実現するために、大手企業間で白熱の競争が繰り広げられる兆しがあるほどです。例えば、世界のスマートフォンの96.4%のオペレーティング・システムを動かす大手企業2社は最近、プライバシー設定の強化を発表しました。ユーザー側のコントロールを追加し、個人のプライバシーを保護するためにデータを暗号化する方向に向かっているのです。⁶³ 世界の最大手ソフトウェア企業はプライバシー保護へのコミットメントを宣伝することで消費者に直接アピールしています。⁶⁴ 顧客の通信内容を解析しないでターゲット広告を実施するなど、企業はプライバシー保護をより強化したサービスやポリシーに変えているほどです。⁶⁵

誤解

自分自身のデータについてのコントロールを失った。

真実

時として、自分のデータについてのコントロールを失ってしまったと見えることがありますが、一方で、それは知ら

米国では、大手自動車メーカーが、自律駐車や衝突回避などの先進機能を実現する、

何百もの

センサーと分析機能を備えた新車の設計を行っています。

米国



ないだけであって、実際には、消費者自身が自己のデータを管理するための様々な有効なツールがあります。例えば、大手ソフトウェア企業の中には、自社が提供するWebブラウザにおいて、訪問するウェブサイトやサードパーティー広告会社が自動的に「Do Not Track (追跡拒否)」要求を受けよう、デフォルトで「Do Not Track (追跡拒否)」機能を設定しています。これにより、追跡されたくない旨をウェブサイトに伝え、Web上で追跡されることからユーザーが保護されます。また、一部のデータブローカーは、ウェブページを設けて、消費者が収集された個人情報を確認し、さらなるデータ収集からオプトアウトしたり、情報の訂正を可能にしています。⁶⁶ これらのツールにより、消費者は自身の情報の収集、使用について好みに応じて制御することができ、特定のマーケティング目的についてのデータ使用をオプトアウトすることが可能です。

データ経済がもたらすインパクト

誤解

データイノベーションは、新規雇用を創出しないばかりか、仕事を奪う可能性がある。

真実

データイノベーションは経済成長を促す強力な原動力になり得ます。実際、米国の61%、欧州で58%の上級役員が、自社の採用計画の中で、データ分析職が重要であると言っています。⁶⁷しかしデータイノベーションはIT業界にのみ雇用を創出するものではありません。1人のデータ関連の雇用が創出されるごとに、3人のIT以外の雇用が創出されると見積もられており、経済全体ではさらに多くの雇用が創出されます。

データイノベーションは、従来のやり方を新しい方法で置き換えるため、職務替えも発生することになりますが、この変化は、インターネット普及の際に、変更されたり廃止された職1件に対し約2.6件の雇用が創出された減少と似ています。⁶⁸大きく雇用が伸びると見込まれる領域の一つがデータアナリティクスです。データから得られる知見を理解することは、しばしば人間の創意工夫によってのみ可能です。今日、データに意味づけをすることができる熟練のアナリストやデータ管理者は世界的に不足しています。McKinsey(マッキンゼー)によれば、米国単独でも、データが秘める可能性を引き出すために必要な、深い洞察力を有する人材が140,000人から190,000人も不足しており、データ分析の内容を理解し、それに基づき意思決定する能力を有するマネージャーやアナリストについては150万人不足していると言われています。⁶⁹

データサイエンティストの2014年の平均基本給は120,000ドル、マネージャーは160,000ドルです。⁷⁰データ駆動型革命がもたらす可能性を存分に引き出すために、世界の企業や政策立案者は迫りくる人材不足解消のために適切な措置を取る必要があると多くの人が示唆しています。

ブラジルにおける、

大手大豆生産企業

の社はソフトウェアとデータアナリティクスに投資することで
ダメージコントロール技術の効率性向上、コスト削減、耕作技術
の増強を図っています。

ブラジル



誤解

データ分析の目的は、人間による判断を取り除くことである。

真実

データにより答えを導き出せるもの（例：町の人口が増えているのか、減っているのか？）があるものの、多くの場合、真相をとらえる答えの多くはそれほど明白ではありません。人は、多様な要素群の関連性に初めから気づいているとは限りません。尋ねるべき問題がはっきりと事前にわかっているとも限らないため、データ分析は質問を繰り返した上で答えにたどり着くという反復プロセスをとることがあります。これらの理由から、相違点の見極めや潜在的に矛盾する点を整理するうえで人間による判断や入力を不要とすることはできません。

データ単体では決して万能ではなく、そして奇跡を起こすことはできないのです。実際、データ自体は、多くの場合が小さな値でしかありません。しばしば乱雑な状態で、元々は別に整理されておらず、また、きちんと構造化されているわけではありません。そのデータを理解し、そして内の関連性を見出すことが難解な作業と言えます。データが問題点の解決に貢献できるかどうかは、より速く素晴らしいソリューションに導くスマートデータストラテジーを使いこなせるかにかかっています。そして、問題点に対する適切な質問を正確に尋ねているかにも因ります。私たちが正しい方法でデータを活用していれば、社会の近々に差し迫った課題のいくつかに対する解決策を明らかにし、革新を焚きつけ、そしてITによって駆動する仕事や経済成長をまた新たに大きく促進することが可能です。

誤解

データイノベーションは大企業のためであり、中小企業には関係がない。

真実

データがさらにユビキタスに広がり、ストレージ費用が下がり、そして強力になったアナリティクスツールが手頃な値段で手に入るようになったことで、中小企業も、かつて大企業のみ利用可能だった高度なデータアナリティクスを存分に活用できるようになりました。たとえば、Intuit社のQuickBooks Onlineで利用可能な「Trends(トレンド)」機能からは、中小企業も同社のユーザーから得られる集合知の恩恵を受け、自社の歳入・支出を他社と比較することで、事業機会を捉えることが可能です。また、経営についてより賢明な意思決定をすることが可能になります。ビジネスインテリジェンスとアナリティクスソリューションの活用は、まだ中小企業の間でそれほど普及していませんが、導入は急速に増えると考えられます。⁷¹ 最近の調査では、中小企業の60%がデータアナリティクスは重要であると考えており、⁷² これは、従業員50名以下の米国企業の57%と、同規模の欧州企業の62%の上級意思決定者による回答です。中堅企業（従業員51名から500名）では、米国の87%、欧州の79%のエグゼクティブがデータアナリティクスは重要であると考えています。

感染パターン

を研究する科学者は、ケニアの携帯電話データを活用して感染発生のホットスポットを特定し、政府に対して

病気根絶の対策

を促すことができました。

ケニア



誤解

データはITセクターにのみ利益をもたらし、その他の経済セクターに恩恵がない。

真実

データは今や技術転換の根幹に位置付けられており、産業界のあらゆる業界において変化と改善をもたらしています。実際、多くの人がデータイノベーションは産業界全体の生産性を向上させると考えています。データ指向の意思決定を行う企業は生産性が5-6%向上したと報告していますが、たとえ、データイノベーションが効率性をわずか1%しか向上させなかったとしても、エネルギーの創出、燃料の節約、低コストでの健康回帰、そして資産の性能、寿命向上といった面で経済セクター全体に波及する影響度は相当のものになります。⁷³ 民間航空産業を例に挙げると、リアルタイムデータを取得することでエンジン効率を向上させ、旅客機の航路を効果的に選択できるようになったことによる、燃料費わずか1%の改善が15年に渡り300億ドルの節約につながるとGEは見積もっています。⁷⁴

誤解

相関性は常に因果関係により導き出される。

真実

「unknown unknowns (未知の未知)」の探索は、データから得られる重要な知見の一つです。しかし相関性に全く意味を見出せないこともあり、相関性が常に因果関係を示すとは限りません。例えば、アイスクリームの売上が急激に上がった時に暴力犯罪や殺人の発生率が急増していたことがあるとしても、アイスクリームを買うという行為が人を殺人犯に仕立てるとするのはあり得ない話です。⁷⁵ それでも、データに因果関係を見つけられるということは、データから見出すことができる貴重な知見の一つであると言えます。そして、興味深い相関関係と因果関係の多くが、日々増加する速さで発見され重要度を増しています。偶然的相関関係と因果関係を見分ける能力は、因果関係と偶然的相関関係を仕分け、信頼度の足らない結果を除去する方法を知る有能なデータサイエンティストがデータ活用に必要なとされる理由の一つです。

誤解

データイノベーションには莫大な予算が必要である。

真実

データ活用には必ずしも莫大な予算を必要としません。多くの場合、少額投資で済み、それでも十分な利益が得られます。例えば、データに対してより包括的なアプローチを採用する企業は、より大きな利益を得ることが可能—この先わずか4年間で、売上増、費用削減、生産性向上から1.6兆ドルに上る「データ配当」を得られます。⁷⁶ 実際、ほとんどのケースでは、小さなデータセット、または既に手元にあるデータを活用することが有効な出発点となり得ます。一週間分の金融取引データについて不正調査をする方が、過去五年間の履歴データを調査するよりはるかに容易に不正行為の傾向を明らかにできます。ある分析によると、データ経済アナリティクスを効果的に活用した企業は、同じ業界の競合他社と比較して26%多くの利益を上げ、社員と有形資産の運用を通じて得た売上は9%多く、市場評価額も12%高いという結果が出ています。⁷⁷

アラブ首長国連邦では、最新のデータツールを使用して、
世界初の

ポジティブエネルギー ビルディング

を設計中です。
消費するより多くのエネルギーを生産する建物です。

アラブ首長国連邦



データの信頼性

誤解

データから得た知見はそのまま常により正確である。

真実

データから得られる知見というのは、基礎となるデータが正確だけでなく、使用されるツールが賢く、そして有能なデータサイエンティストによる偏った結果の算出を避けるモデルの開発があって初めて、正しいと言えます。例えば、ボストンで開発された革新的なアプリ「Street Bump」は、加速度計とGPS機能を導入しているスマートフォンを活用して、路上でユーザーがデコボコに乗り上げる度にデータを収集し道に凹凸のある場所を調査、特定します。しかし、もしスマートフォンユーザーに若い世代が多いと、アプリからのデータは町の全人口層から凹凸データを正確に取得しているとは言えず、町の凹凸への対処は偏ったものになる危険性があります。だからこそ、統計学的関連性を確実に発見しデータの結果に偏りが生まれないように、データサイエンティストが関わる必要があるのです。

誤解

自身の直観による決定の方が、正しい決定であることが多い。

真実

人は、生まれてからというもの、意思決定に自身の直感を信ずることが身についています。実際、世界の経営者のうち約19%が自身が本能と勘にほぼ全面的に依存する「直感的意思決定者」であると評しています。⁷⁸しかし今日、データがより高い精度と高速処理能力、そして大きな影響力を備えるようになり、情報に基づく意思決定を支援できるようになりました。業界全体のIT意思決定者を対象とした調査によると、内59%が、データ技術に投資を促進する第一の目標は意思決定における質の向上であると回答しています。⁷⁹これがあてはまるのは職場だけではなく、私たちが日々の生活の中でも直感的な意思決定をしているからです。いくつかの例によると、私たちはより速く正しい選択を可能にするツールから多大な恩恵を受けることができます。例えば、ヒューマンエラーは自動車衝突事故のおよそ93%を占める犯人ですが、周囲の状況に関す

る膨大な量のデータを収集する新しい車両センサーにリアルタイム分析と自動意思決定（完全自律式ではなく）を組み合わせるにより自動車関連の死傷者は50%ほど減少したという事例があります。⁸⁰

誤解

データが大きければ大きいほど、優れたデータである。

真実

優れたデータセットは、必ずしも大きいとは限りません。実際、大きさは最も重要度が低い要素の一つかもしれません。適切なツールを用いて分析すれば、たとえわずかな量のデータでも、多くの問題、疑問に対する鋭い知見が導き出され得ます。これらの知見が得られるかどうかは、基礎となるデータの質と、使用するツールの質に大きく依存します。データが多ければ多いほど、より多くの真実が得られ、客観性が高まると考えられる場合もあります。しかし、例えば、地元の店の閉店時間、オイルを最後に交換してから走行距離、今月の電話料金などの様に、簡単なデータで即座に問題が解決する場合もあります。問題によっては、ウェブページや、Excelスプレッドシート、CRMデータベース内に含まれるデータ量が小さくても、ビッグデータと同様に問題解決に大きく貢献できる場合もあるということです。最も重要なのは、問題解決が必要な時に、場面を問わず活用できるよう、サイズに関係なく安定したデータが生成され、安全に蓄積されており、アクセス、処理が可能になっていることです。

気象データ

は単に傘の携帯が必要か予測するのに使われるだけではなく、作物の産出量やインフルエンザの流行の予測に活用されたり、モロッコにおいて必要な脱塩能力を見積もるために役立てられています。

モロッコ



誤解

構造化されていないデータに利用価値はない。

真実

構造化されていないデータも役立つ場合があります。実際、非常に有益なデータの中には、列と行の形に整然と収まったものとは限りません。テキスト文書やX線画像のように非構造化である場合もあります。重要なことは、データを必要とされる最も便利な形式にデータを変換することです。フォーチュン1,000企業は、自社の有するデータの有用性をわずか10%高めるだけで、20億ドル以上の増益を見込める可能性があります。⁸¹

誤解

データは収集された本来の目的のためにのみ使用される。

真実

クラウド技術が、データストレージの価格を劇的に低下させたことにより、一度保存したデータをストレージ容量確保のために削除するということは必要なくなりました。つまり、他の資源と異なり、データは一度使ったら終わりの使い切り資源ではないのです。データは、しばしば、他のデータセットと結合できる再生可能資源となり、最初にデータを作成した時には予期されなかった問題に対して解決策を導き出すために何回も使用されることがあります。例えば、気象データは単に傘が必要かどうか判断するのに使われるだけではなく、作物の収穫量やインフルエンザの流行予測に活用されたり、モロッコにおいて必要な脱塩能力を見積もるために活用されたりします。

21世紀のデータ革新

誤解

データから莫大な利益が得られるのは、まだはるか先の未来である。

真実

データイノベーションが数十年先の将来に渡り利益を生み出し続けるとみられる一方、身の回りの至る所で、既に大きな利益がもたらされています。毎日テラバイト単位のデータが既に活用されており、正確な天気予報がなされていることはもちろん、何百万ものGPS信号から入手したデータが目的地への移動時間を予想し、ナビゲートしてくれます。また、何百万もの診断データを活用することで病気の原因を特定したり、テラバイト単位の金融報を活用してクレジットカード詐欺を未然に防止することもできます。

実際、エコノミストの調査部が、調査回答者に過去五年間で組織に影響を与えたデータについて尋ねたところ、約10%がビジネスの仕方が根本的に変わったと回答し、⁸² 回答者の46%が、ビジネス上の意思決定を促す重要な要因となったと述べています。新しい調査によると、今後4年間だけで、データが新しい機会とソリューションをもたらし、データを最適に活用することで、1.6兆ドルに上る「データ配当」が得られると見積もっています。⁸³ また、本年度行われた別の調査では、米国の33%、欧州の24%の上級役員が、自社の成長の10%以上がデータアナリティクスに関係していると考えています。⁸⁴ 今後五年先の見通しについては、米国の58%、欧州の43%の上級役員が同様の予測をしています。

2004年の南アジア大津波が発生した後、インドネシアの漁師たちに携帯電話が与えられました。これにより彼らの収入は**30%**上昇しました。魚の実市場価格に関するデータを初めて手に入れることができるようになったからです。

インドネシア



誤解

データはただの幻想に過ぎない。

真実

データを活用して人が問題を解決するということは、決して新しいことではありません。私たちは、近代文明以来、食糧を供給する農業技術、世界貿易を推進する航行技術、コレラによる何百万人という死者を出さないための健康に関する知識を得るためなど、多くの優れた意思決定のためにデータを活用してきました。しかし、これまで、データは保管に費用がかかり、利用が難しい、限定的な資源に過ぎませんでした。今日、データは豊富に得られ、ストレージコストは低下し、強力なツールにより効果的な利用が可能になりました。その結果、一連の新たな課題に直面しても、強力な新しいデータアナリティクス技術のおかげで、増え続ける膨大な量のデータを上手く分類し、非常に難解な課題に対しても、効果的な知見を発見し、思いもかけないソリューションを得ることができるのです。

誤解

IT 駆動型経済成長の時代は終わり、データイノベーションが生産性を高めることはない。

真実

ITによるイノベーションは、主にそれがもたらす生産性を向上させる力により、経済を成長させ、雇用を創出し、世界中で生活水準を高めました。例えば、1990年代を通して1-2%ほど生産性が向上しました。⁸⁵ ノースウェスタン大学の著名な経済学者は、ITによるイノベーションから多くの利益が得られた時代は既に過ぎ去ったと主張しています。⁸⁶ しかし、「生産性の時代」は未だ健在です。実際、データ指向型の意思決定を行う企業は生産性が5-6%向上したと報告しています。⁸⁷ 未だ展開途上のデータ活用機会が米国における生産性を仮に20年間で年率1.5%向上させるだけでも、平均国民所得を30%引き上げるに十分足る費用効果を得ることができます。⁸⁸ 効果的にデータを活用することで、産業界全体の効率性をわずか1%でも高めることに成功すれば、2030年までに全世界のGDPに対し約15兆ドルもの経済効果をもたらすと経済学者は予測しています。それは、今のアメリカの経済規模と同等の経済規模が追加されることを意味しています。⁸⁹

バルセロナではデータを活用して

スマート・シティ

の構築、行政サービスの向上、そして

持続可能な交通

ソリューションの提供を行っています。

スペイン



世界にもたらされるデータイノベーションの恩恵

誤解

今日、データから恩恵を受けているのは米国だけである。

真実

世界中で、データは重要な課題解決に既に利用されています。例：

- + 研究者は、毎秒1,000以上のデータポイントを追跡し、関連づけることで、異常に安定した生命徴候を示す未熟児は、その翌日に深刻な発熱を発することを示し、医師たちに衝撃を与えました。これにより、医師が予防処置を取ることができ、救命に繋がりました。
- + ブラジルにおける、大手大豆生産企業の一社はソフトウェアとデータアナリティクスに投資することでダメージコントロール技術の効率性向上、コスト削減、耕作技術の増強を図っています。
- + バルセロナではデータを活用して スマート・シティの構築、行政サービスの向上、そして持続可能な交通ソリューションの提供を行っています。⁹⁰
- + アラブ首長国連邦では、最新のデータツールを使用して、世界初のポジティブエネルギービルディングを設計中です。消費するより多くのエネルギーを生産する建物です。⁹¹
- + インドの農家は、種苗や人工衛星、センサーやトラクターの情報を活用し、作物の種類や植え付け方法、農場から地元民に届くまでの鮮度の追跡方法、そして気候変化への適応方法について、より効果的な意思決定を行っています。⁹²

誤解

新興国はデータ分析を優位に活用する環境が整っていない。

真実

データ革命とそれがもたらす恩恵は世界的な現象です。データによる最も重要な恩恵と機会は、技術がしばしば遅れがちな新興国にもたらされます。IDCは、デジタル分野において新興国の成長率は、2012年から2020年の間で36%から62%に急成長を遂げており、2017年までに成熟国を凌ぐと予測しています。⁹³ 同様に、新興国のNGOを対象に行われた調査では、90%がデータアナリティクスが最も重要なツールとして、現地の人々の支援に繋がる優れた知見をもたらすだろうと回答していました。⁹⁴

以下は、新興国で挙げられるデータイノベーションの事例です。

- + ケニヤでは、モバイルデータを使用することでマラリアの感染パターンを解析し、政府による撲滅活動を最適化するための場所特定に役立っています。⁹⁵
- + 2004年に南アジア大津波が発生した後、インドネシアの猟師たちに携帯電話が与えられました。これにより彼らの収入は30%上昇しました。魚の実市場価格に関するデータを初めて手に入れることができるようになったからです。⁹⁶
- + ペルーでは、遺跡が開発による脅威にさらされています。政府は、航空技術と、イメージを結び付ける強力なソフトウェアを駆使し、3次元データポイントクラウドを完成させました。これを活用し、危機に瀕している自国の貴重な宝をマッピング、監視、保護しています。⁹⁷
- + 文化的遺産を保護するために、ベトナム政府は3Dスキャナーを使用して40,000件の歴史資産を今後5年に渡ってデジタル化します。⁹⁸
- + インドでは、インターネット、キオスクが4百万以上の農民に、作物の価格、天候、その他の情報を現地語で提供しています。高度なアナリティクスやモバイル技術を駆使して、個人農家から得られるデータを追跡、分析することで、肥料や種苗などの必要なものを農民に供給することができるのです。⁹⁹

文化的遺産を保護するために、ベトナム政

府は 3D スキャナーを使用して

40,000 件の 歴史資産

を今後 5 年に渡ってデジタル化します。

ベトナム



データの規制における政府の 役割

誤解

データは、低所得者や恵まれない地域が直面している不利益をさらに強めるツールとして使われる。

真実

一部の人や特定の層に不公平な方法でデータが使用されているという懸念を解消することは、業界や政府が最優先で対応すべき課題です。同時に、不当な差別と戦い、人々に力を与えるためにデータを活用する機会を見逃してはなりません。責任ある使い方をすることにより、データは強力な新しいツールとして、職、資産、教育などへのアクセスの機会を阻む差別を解消する力となり得ます。¹⁰⁰

誤解

政府には果たすべき役割がない。

真実

遠隔で実行できるコンピューティング能力、ストレージ、アナリティクス、ソフトウェアサービスの急成長により、様々な新しい政策課題が浮上しています。多くの条約、法律、規制は、策定時点では、これらの可能性を予見していませんでした。この状況は、政府そしてデータ経済の成長にとって問題です。プライバシーとセキュリティのバランス、自由貿易、基礎研究、そして労働力確保といった長期的政策課題について、これらの技術が爆発的に成長していることを受けて、改めて検討する必要があります。

政策立案者が、データ革命の可能性を最大限引き出し、すぐそこまで来ている生産性や経済の成長、消費者の利益の新たな波を加速させるためにとれる具体的なステップがあります。教育、食糧生産、生活の確保、エネルギー節約、交通、経済成長といった社会が抱える最大の課題の解決に向けてデータソリューションの進化に向けて投資することができます。国際的な法順守に関する連携において、明確なルールを確立することができます。テクノロジー業界における、投資、イノベーション、取引、成長する能力に影響がある市場の自由度を促進することができます。物理的境界が妨げられないデータの自由流通を確保することができます。有能な人材を確保して、近々に起こり得るデータサイエンティスト不足の課題を解決することができます。そして、イノベーションを活性化し、データから得られる恩恵が企業と消費者両方にもたらされるよう促すことができます。政策立案者は、人の才能と努力を活用するために有効な政策をもって、イノベーションと投資を解き放ち、量的及び質的に能力を拡大し、データ中心型経済がもたらす恩恵を最大限に引き出すことができます。

ペルーでは、遺跡が開発による脅威にさらされています。政府は、航空技術と、イメージを結び付ける強力なソフトウェアを駆使し、

ペルー



3次元

データポイントクラウド

を完成させました。これを活用し、危機に瀕している自国の貴重な宝をマッピング、監視、保護しています。

誤解

データ・ローカライゼーションが、プライバシー保護とセキュリティ向上に有効である。

真実

データ・ローカライゼーションは、データを自国の国境内に蓄積させるため、プライバシーとセキュリティを向上させると信じている人がいます。しかし、今日の技術の恩恵は、インターネットという世界的な力によってもたらされ、その促進は異なるデータセンター間を自由に流通するデータによって実現するものです。国境を越えて流通するインターネットデータ量は、2005年から50%以上増加しています。¹⁰¹ 国境を越えて自由にデータが往来できるようになったため、中小企業や起業家は、世界中に店舗を構えることができ、国をまたいで商品、サービスやアイデアを即時に手に入れ、販売を開始することができるようになったのです。しかし、どこの国の政府も、しばしば、データの自由な流通を制限したり、市場参入の条件として自国内にデータサーバーを設置するよう要求する政策を検討することあります。¹⁰² こうした制限は、データイノベーションから得られる非常に大きなスケールメリット、経済的な利益、そして、異なる場所に存在する様々なデータセットを組み合わせたデータから有益な知見を得る能力を弱体化させます。また、自然災害や技術的障害が発生した時に対処するために、貴重なデータを他の地域においてバックアップしておく行為を妨げ、セキュリティまでが弱体化してしまいます。データがもたらす利益を手に入れるためには、すべての国の法律を同一にする必要はなく、その互換性が求められています。国境を越えてデータが自由に往来できるようにすることは、データ駆動型の恩恵を実現するための基本原則の一つです。

誤解

データが保護され得る唯一の方法は、政府が介入し、データの保護を要求することである。

真実

政府が施行する現行のプライバシー規則と、厳格かつ革新的なプライバシー先進技術およびデータを安全に管理し個人情報保護するための業界の自主規制とを組み合わせることは可能です。これと反対に、自国内にデータを蓄積するよう要求するといった方法で、個人のプライバシーやセキュリティの問題に対処しようとする政府の命令は、革新を阻害し、データイノベーションが社会にもたらす利益を限定してしまう可能性があります。



デジタル用語集

データの用語を理解する

豊富なデータ

全く新しい方法で有効な形式のデジタルデータを収集することが可能になったこと、データの保管コストが下落したこと、データから価値を見出す新しい方法が得られたことで、かつて少なかったデータが、今日では豊富に利用可能となりました。

アダプティブ・インテリジェンス

アダプティブ・インテリジェンスとは、単純にデータ統計処理を行うのではなく、それを特定領域における専門知識を含むデータと組み合わせて得られるコンピューター・インテリジェンスを指します。専門知識と行動モデルに関するインテリジェンスを組み合わせることにより、システムは事象からさらに多くを学び、多様な状況に適応できるようになります。

アルゴリズム

アルゴリズムとは、数学的手法を用いてデータを分析し、問題を解決する手順、または一連のコンピュータ命令を指します。アルゴリズムは、ほとんど全てのソフトウェア・プログラムで使用されています。

アンビエント・インテリジェンス

アンビエント・インテリジェンスとは、環境知能のことを言い、コンピューティングがますます低コストでユビキタスになりテクノロジーが環境に溶け込むことで見えなくなり、ほぼすべてのものが知的になり、事象に反応し、つながっていることで必要な時にいつでも利用できるという将来ビジョンを指しています。

アナリティクス

アナリティクスは、データの中から重要な知見、パターン、関連性を見つけることを目的に、統計手法とソフトウェアのアルゴリズムを同時に適用することを言います。

異常検出

異常検出とは、予想されるパターンと一致しないデータセット内のデータ項目を特定することを言います。異常は、データ内の外れ値、例外、または混入物とも呼ばれ、しばしば、重要かつ有益な情報を提供することがあります。

匿名化

データの匿名化とは、個人の身元特定につながる可能性のあるあらゆる個人情報の除去する処理を言います。

不良データ

不良データは、欠落のある、または不正確なデータを指します。住所の間違いなど単純なことも該当しますが、不良データはフォーチュン1,000に名を連ねる企業に、毎年数十億ドルもの損害を与えています。

ビッグデータ

ビッグデータは、膨大な量のデータに対してコンピューターによる分析を適用するプロセスを指す包括的な用語で、新しい知見の発見や意思決定向上のために使用されます。多くの場合、データ量が非常に大きく、種類が多様で、動きが速いため、従来のデータ処理ツールを使用して処理することが難しいデータセットのことを指します。

ブロントバイト

ブロントバイトは、非常に膨大な量のデータを指す単位です。1ブロントバイトは、一般的に、1,000 ヨタバイトに相当し、1の後にゼロが27桁続く数字です。



ビジネスインテリジェンス (BI)

ビジネスインテリジェンス (BI) とは、生データを、ビジネス・パフォーマンスや意思決定向上につなげる経営知識に変換させる技術やアプリケーションのセットを指します。

クラスタ分析

クラスタ分析は、互いに類似しているデータオブジェクトを特定し、それらを集合化する処理で、データ間の相違・類似性を深く理解するために使用されます。

クラウド

クラウドは、リモートでホストされているあらゆるアプリケーション、サービス、データを参照する広義の用語です。一般的には、ネットワークアクセスを通じてコンピューティング又はストレージリソースに対するユビキタスかつオンデマンドなアクセスを可能にする相互にネットワーク接続された大規模なリモートサーバ群のことを指します。

コグニティブコンピューティング

コグニティブコンピューティングは、膨大な量の情報と、機械学習技術、パターン認識技術、そして時には人間の脳が機能する方法を模倣する自然言語処理とを組み合わせる処理を指します。これらのシステムは、情報源に対して文脈と知見を組み合わせることによって、学習し、そして人間と対話することが可能です。

コンピューター生成データ

コンピューター生成データとは、人の介入なしにコンピューターによって自動生成されるデータを指し、例えば産業機械から入手するコンピューターログファイルや人工衛星の位置データ、センサーデータなどを言います。

ダークデータ

ダークデータとは、保存されているが、分析処理されていない、非構造型の未活用データで構成され、放置されていると考えられるデータを指します。

データ

データとは、条件、物品、またはアイデアを表現するためにデジタル形式で操作処理が可能な加工前の整理されていない形式の情報を言います。データの一般的なタイプは、売上高、市場調査結果、気象センサーの計測値や、都市、人口のリストなどがあります。私たちは今日、約25京バイト／日を生成しています。

データ集約

データ集約とは、より高度な分析を行う目的で複数のソースからデータを収集する行為を言います。

データ集約ツール

データ集約ツールとは、複数のソースから入手した散在するデータを単一の新しいデータセットに変換する目的で使用するツールを指します。

データアナリティクス

データアナリティクスとは、データを変換、モデル化することにより有益な情報、知識、有意性を派生させることを目的として使用するソフトウェアアプリケーションのことを言います。データに秘められたパターンや未知の相関性を明らかにして、意思決定を支援するためによく使用されます。

データアナリスト

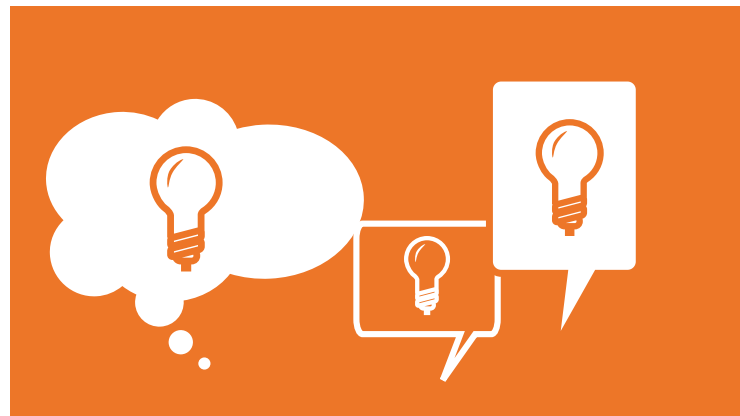
データアナリストとは、データ準備、整理、処理の責任者のことを指します。

データ・アーキテクチャおよび設計

データ・アーキテクチャの作成は、一般に、データが処理、蓄積、使用、アクセスされる方法を設計、構造化するために、新システムの検討段階で行われます。システム内においてデータを適切に保護する為に、特定のデータがどのように相互に関連し、処理されるかを定義することで、データの往來の管理方法を設計することを指します。

データベース

データベースとは、大規模な構造化されたデジタルデータのセットを指し、格納されたデータを迅速に検索、アクセス、更新することができるよう設計されたものです。



データセンター

データセンターとは、ネットワーク上のサーバーとデータストレージリポジトリを収容する物理的な施設で、一般的にリモートアクセスが可能な莫大な量のデータを遠隔で蓄積、処理する際に使用されます。推定50万のデータセンターが、世界中に設置され、その多くがクラウドを構成しています。

データクレンジング/クリーニング

データクレンジングとは、生データを検証し、是正する処理のことを指し、重複の検出・削除、エラーの訂正、欠落データの追加、破損データの除去、等を通して、一貫性を向上させます。

データ指向型の意味決定

データ指向型の意味決定を使用する企業は、データを収集、処理、分析し、重要な意思決定の補助に活用します。マサチューセッツ工科大学・スローン経営大学院のエコノミストであるエリック・ブリニョルフソン氏による研究によると、データ指向型の意味決定を使用する企業は、生産性が5～6%高まることがわかっています。

データマイニング

データマイニングは、強力なコンピュータ・アルゴリズムを使用し、大規模なデータセットの中からパターンや知識を見い出します。

データ品質

データ品質とは、ユーザーにデータの価値を測るために使用される指標のことです。意思決定、計画、または経営に必要なデータの信頼性、効率性、および有用性と関連します。

データ科学

データ科学とは、複雑な問題を解決することができる重要な知見を抽出することを目的とした、統計データの可視化、コンピュータプログラミング、データマイニング、機械学習、及びデータベース技術を組み込んだ領域のことを言います。

データサイエンティスト

データサイエンティストとは、コンピュータアルゴリズムを開発、活用することで、例えば、人間の知識、数学的なノウハウ、技術ツールを組み合わせデータを理解する能力を有する者を指します。

データセキュリティ

データセキュリティとはデータを破壊、誤用、不正アクセスから保護する方法です。適切なデータセキュリティ対策をすることにより、データの漏洩を防止し、データの整合性を確保し、プライバシーを保護することができます。多くの場合、人、プロセス、技術を集中的に組み合わせることが必要です。

データセット

データセットとは、関連する一連の情報の集合を指し、典型的には一つのユニットとして操作することができる表形式を構成する要素群を指します

データソース

データソースとは、主となるデータの発生源を指し、例えば、データベース、スプレッドシート、またはデータストリームなどがあります。

データの可視化

データの可視化とは、より効率的にデータの意味を抽出したり、情報を伝達するためにデータを視覚化することを指します。

データの仮想化

データの仮想化とは、データがどこに存在し、またはどのように構成されているかなど、技術的な詳細情報を知る必要なしに、様々なデータソースを取得し、操作するためのプロセスを指します。

非識別化

データの非識別化とは、個人と特定の情報を関連付ける情報を除去するための処理を言います。



破壊的変革

破壊的変革とは、社会、ビジネスにおける大きく抜根的な変化を指し、働き方、日常生活、遊び、価値の創造について全く新しいコンテキストをもたらし本質的に新しい技術により実現されます。データイノベーションは、しばしば、破壊的変革を起こす技術として扱われます。

エグザバイト

エグザバイトとは、1のあとにゼロが18個続く数字で、莫大な量のデータストレージの容量を表す単位として使用されます。分かりやすく説明すると、今日、私たちは毎日1エクサバイトの新しい情報を生成しています。

Hadoop

Hadoopとは、分散ファイルシステム間でやりとりされる膨大な量のデータを処理、蓄積するために構築されたオープンソースのソフトウェアフレームワークです。

モノのインターネット (IoT)

モノのインターネットとは、身の回りの機器が賢くなり、インターネットとつながり、私たちの指先から、取り巻く環境全体まで、スマート革命が広がる世界を表します。接続可能なあらゆるものが接続されるため、より適切にInternet of Everythingとも表現されます。一説では、インターネットに接続が可能な世界中のもののうち、わずか1%しか接続できていないとされています。2020年までに、およそ500億の機器がインターネットに接続されると予測されています。

レガシーシステム

レガシーシステムとは、旧式で廃れたコンピューター、アプリケーション、技術でありながら、一方で必要な機能を実行しているために使用され続けているものを指します。

機械学習

機械学習とは、特定のパターンや事象が発生した時に、とるべき行動について、経験から「学ぶ」ことを目的に、コンピューターにデータを分析させるアルゴリズムの使用法を指します。

メタデータ

メタデータは、データに関するデータです。これには、データ作成者、作成日時、ファイルサイズ、最終更新日などデータに関する基本的な概要情報が含まれます。

異常値検出

異常値とは、大きなデータセット内の通常平均から大きく逸脱したデータのことを指します。他のデータからは数値的に大きく離れているため、異常値の存在は、何かが起きていることを示しており、したがって、更なる分析が必要であると言えます。（「異常検出」を参照。）

パターン認識

パターン認識とは、データ内のパターンを検索、識別するプロセスです。DNA配列内の配列セットの繰り返しを識別するといった単純なパターン認識もあれば、2つのデータセットが互いに連関するパターンを見つけることにより、一つの出来事が他の出来事と関連性を有するか検証する場合もあり、他にも、機械学習によって画像内に含まれる数字を探すような複雑なパターンを探すことを指す場合もあります。

ペタバイト

ペタバイトとは1のあとにゼロが15個続く数字で、ストレージの容量を表わす単位として使用されます。1ペタバイトは100ギガバイトに相当し、米国議会図書館に所蔵されるデータの約4倍の量にあたります。

予測分析

予測分析とは、1つ以上のデータセットにソフトウェアアルゴリズムを使用することで将来の出来事や傾向を予測することを指します。現在のデータが過去と比較できるのであれば、しばしば、将来の予測に使用することができます。

予測モデリング

予測モデリングとは、傾向や将来の行動、結果の予測に使用されるモデルを開発するためのプロセスを言い、多くの場合、現在と過去の事象を比較することで行います。



リアルタイムデータ

リアルタイムデータとは、作成されたと同時に使用されるデータを指します。多くの場合、作成、加工、蓄積、分析までがミリ秒以内で実行されます。リアルタイムデータは、株式市場の価格から、自動車のアンチロックブレーキシステムで使用される車輪速度まであらゆるデータを含みます。

リコメンデーションエンジン

リコメンデーションエンジンとは、機械学習技術を介して導出された様々なデータパターンに基づいて、ユーザーに対して、推薦、提案その他パーソナライズされた何かを提供するためのコンピュータアルゴリズムを指します。

回帰分析

回帰分析とは、二つ以上の変数の間の関係を推定するためにデータを使用するための統計的処理手法の一つです。

リスク分析

リスク分析とは、ソフトウェア、データ分析ツールを使用して、プロジェクト、アクション、または意思決定についてあり得るリスクを識別することです。最新のデータツールは、そのようなリスクを特定し、組織が直面するリスク軽減につながる一連のシナリオのモデル化を図り、システムを監視することで、実行と計画が乖離し始めた場合に問題特定を助けます。

根本原因分析

根本原因分析とは、障害や問題の根本原因を特定するために原因と結果の関係性に焦点を当てて調査し問題解決を図る手法を指します。その原因が取り除かれることにより、望ましくないできごとと反復発生が防止されるのであれば、それが根本的な原因であったといえます。

半構造化データ

半構造化データとは、データベースで使用されるほどの正式なデータモデルを用いて構成はされていないが、データと階層を記述する何らかの方法を提供するものを言います。半構造化データは、しばしば、タグや他のデータマーカーを使用し、自己記述型といわれる構造を有します。

スモールデータ

スモールデータとは、顧客調査に含まれているような、少量のデータであっても、結果を生み出す行動につなげるために活用することを言います。一般に、人間が理解し分析することができる程の小さなサイズのデータを指します。

構造化データ

構造化データとは高度に整理され、多くの場合、行と列で構成された検索や操作が容易なデータを指します。

テラバイト

テラバイトとは、12桁以上のデータ量を表す単位として使用されます。1テラバイトのハードドライブは、家庭や職場のコンピューターのストレージ容量としてよく見られるようになり、クラウドを介してアクセスすることもできるようになりました。分かりやすく表現すると、1テラバイトは、HDビデオ約300時間分のデータ量に相当します。

テキスト分析

テキスト分析はデータから有意性の導出や概念の抽出、物事の実態を見抜くため、テキストベースのデータに統計的・言語的、および機械学習技術を使用することを言います。テキスト分析は、一般書、転写物、ウェブ投稿、解説、またはフォームに含まれるような自然言語テキストで実行されます。これはコンテンツの要約、発見、または分類に役立ちます。

取引データ

商品購入、請求、支払い、および出荷などの取引行為に関するデータから導出されたデータです。これらは通常、時刻が紐づいており、組織における日常の業務をサポートします。

非構造化データ

非構造化データとは事前に定義された構造を持たず、例えば、会議の議事録のようなデータを指します。ある推計によると、組織内のデータ全体の70%から80%以上を非構造化データが占めていると言われています。



Variety (種類)

データイノベーションを定義する四つのVの一つである「variety (種類)」とは、合成・分析して知見を得るために様々なソースから入手する多様な種類のデータのことを指します。今日、アプリケーションで処理されている様々なデータには文字のデータベース、取引データ、ストリーミングデータ、画像、音声、動画が含まれます。

Velocity (速度)

データイノベーションを定義する四つのVの一つである「velocity (速度)」とは、データが作成、蓄積、分析、可視化される速度のことを指します。例えば、大規模なデータウェアハウスは、毎日数十億行に上る新しい情報を受け取ります。即時性を要求されるデータについて、その価値を最大化するためには、情報が流れ込むと同時に並行的に使用されなければなりません。

Veracity (正確性)

データイノベーションを定義する四つのVの一つである「veracity (正確性)」とは、データの正確性、確実性、精度を示すために使用されます。

Volume (量)

データイノベーションを定義する四つのVの一つである「volume (量)」とは、メガバイトからペタバイトに渡るまで処理されるデータの量を表わします。

ヨタバイト

ヨタバイトとは1のあとにゼロが24個続く数字で、非常に大きなデータストレージの容量を表す単位として使用されます。分かりやすく説明すると、1ヨタバイトは2500兆枚のDVDに保存されたデータ量を表します。

ゼタバイト

ゼタバイトとは1のあとにゼロが21個続く数字で、ストレージの容量を表す単位として使用されます。2013年の時点で、世界のウェブ(WWW)上のデータ規模は、約4ゼタバイトに達したと言われています。2016年までに、毎日1ゼタバイト以上のデータが世界中のネットワークを流通するすると予測されています。

後注

- ¹ Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making*. Cap Gemini, 2012. ウェブサイト POV: <https://www.capgemini.com/resources/the-deciding-factor-big-data-decision-making>
- ² A 1 percent productivity increase may seem small, but as Jeff Immelt, CEO of GE puts it, “tell an oil guy you can use software to save him one percent on something, and that guy will be your friend for life.” Evans, Peter C., Marco Annunziata. *Pushing the Boundaries of Minds and Machines*. GE, 2012. ウェブサイト <http://files.gereports.com/wp-content/uploads/2012/11/ge-industrial-internet-vision-paper.pdf>
- ³ An estimated 50 billion devices will be connected to the Internet by 2020 enabling the Internet of Everything. Evans, Dave. *The Internet Of Things How The Next Evolution Of The Internet Is Changing Everything*. Cisco Internet Business Solutions Group (IBSG), 2011. ウェブサイト: ホワイト・ペーパー: http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf
- ⁴ “Bringing big data to the enterprise.” IBM. <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>
- ⁵ Golub, Ben. “Enough Data to Fill a Stack of DVDs to the Moon (and Back).” *ComputerWorld* 2011. ウェブサイト: http://blogs.computerworld.com/18351/a_stack_of_dvds_to_the_moon_and_back
- ⁶ Corry, Will. “BIG Data / The Volume Of Business Data Worldwide, Across All Companies, Doubles Every 1.2 Years, According To Estimates.” *The Marketing Blog* 2012. ウェブサイト: <http://www.themarketingblog.co.uk/2012/10/big-data-the-volume-of-business-data-worldwide-across-all-companies-doubles-every-1-2-years-according-to-estimates/>
- ⁷ “Digital Imaging in the Cloud.” *There Magazine* 2012: 16. ウェブサイト: http://www.agfahealthcare.com/he/global/en/binaries/THERE_12_tcm541-95647.pdf
- ⁸ The 787 uses data sensors to reduce fuel, monitor systems, and even employs accelerometers in the nose of the plane to counteract turbulence. If the sensors register a sudden drop, they immediately tell the wing flaps to adjust (in a matter of nanoseconds) and in so doing, what used to be a 9 feet drop in an older plane can be reduced to just 3 feet in the 787, making for a much smoother flight. Humphries, Matthew. “The Boeing 787 produces over 500GB of Data during Every Flight.” *Geek.com* 2013. ウェブサイト: Gosling, Kevin. “E-Enabled Capabilities of the 787 Dreamliner.” *Aero Quarterly* 2009:22-24. http://www.boeing.com/commercial/aeromagazine/articles/qtr_01_09/pdfs/AERO_Q109_article05.pdf
- ⁹ Jet engine maker GE says the engine data allows it to figure out things like possible defects 2,000 times as fast as it could before. Hardy, Quentin. “What Cars Did for Today’s World, Data May Do for Tomorrow’s.” *New York Times* 2014. B7. 印刷物: http://bits.blogs.nytimes.com/2014/08/10/g-e-creates-a-data-lake-for-new-industrial-ecosystem/?_php=true&_type=blogs&_php=true&_type=blogs&module=BlogPost-Title&version=Blog%20Main&contentCollection=Big%20Data&action=Click&pgtype=Blogs®ion=Body&r=1&
- ¹⁰ Henschen, Doug. “Bad Winter Weather Meets Big Data Prediction.” *InformationWeek* 2014. ウェブサイト: <http://www.informationweek.com/big-data/software-platforms/bad-winter-weather-meets-big-data-prediction/d/d-id/1114192>
- ¹¹ Groenfeldt, Tom. “At NYSE, The Data Deluge Overwhelms Traditional Databases.” *Forbes* 2013. ウェブサイト: <http://www.forbes.com/sites/tomgroenfeldt/2013/02/14/at-nyse-the-data-deluge-overwhelms-traditional-databases>
“NYSE Euronext Adapting To Market Changes with Near-Real-Time Insight into Information.” IBM Corporation, 2013. 印刷物: NYSE Euronext. <http://www.ibmbigdatahub.com/sites/default/files/document/NYSE-Euronext-IMC14787USEN.PDF>
- ¹² Mayyasi, Alex. “Why UPS Trucks Don’t Turn Left.” *Priceonomics.com* 2014. ウェブサイト: <http://priceonomics.com/why-ups-trucks-dont-turn-left/>
- ¹³ *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2014–2019 White Paper*. Cisco Visual Networking Index, 2015. 印刷物: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html
- ¹⁴ “LSST General Public FAQs.” LSST.org 2015. <http://www.lsst.org/lsst/faq>
- ¹⁵ Quazi, Ed. “Next Generation Sequencing and the Big Data Bottleneck.” *Arkivum* 2014. ウェブサイト: <http://arkivum.com/next-generation-sequencing-and-the-big-data-bottleneck/>
- ¹⁶ Valerio, Pablo. “Internet Of Things: 50 Billion Is Only The Beginning.” *EE Times* 2014. ウェブサイト: http://www.eetimes.com/document.asp?doc_id=1321229
- ¹⁷ In 1980, there was a rule of thumb that one needed a data administrator for 1GB of storage. At that time a GB of disk cost about a million dollars, and so it made sense to have someone optimizing it and monitoring the use of disk space. Gray, Jim, Prashant Shenoy. “Rules Of Thumb in Data Engineering.” Redmond, WA: Microsoft Research Advanced Technology

- Division, 2009. 印刷物: テクニカルレポート: http://research.microsoft.com/pubs/68636/ms_tr_99_100_rules_of_thumb_in_data_engineering.pdf
- ¹⁸ Wohlsen, Marcus. "Dropbox Slashes Its Price As The Cost Of A Gigabyte Nears Zero." *Wired* 2014. ウェブサイト: <http://www.wired.com/2014/08/dropboxs-plan-to-stay-relevant/>
- ¹⁹ From more than \$200,000 a gigabyte in 1980 (even up to million dollars) to \$0.02 per gigabyte in 2013 Meeker, Mary. "Kleiner Perkins Internet Trends 2014." 2014. プレゼンテーション: <http://cryptome.org/2014/05/internet-trends-2014.pdf>
- ²⁰ Based on average fuel efficiency of passenger cars in 1980 (24.3 mpg), enabling a person to buy 10 million times more capacity for the same price leads to 10 million gallons of gas, which could fuel 243 million miles of travel. If the circumference of the earth is 24,901 miles, then a person would be able to circle the earth 9,758 times or nearly 10,000 times. U.S. Department of Transportation. スライド 4-23: Average Fuel Efficiency Of U.S. Light Duty Vehicles. Washington, DC: Bureau of Transportation Statistics, 2013. 印刷物: http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics/html/table_04_23.html
- ²¹ Savitz, Eric. "Big Data: The Hidden Opportunity." *Forbes* 2012. ウェブサイト: <http://www.forbes.com/sites/ciocentral/2012/05/01/big-data-the-hidden-opportunity/>
- ²² The world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s according to research by Martin Hilbert and Priscila López. Hilbert, M., P. Lopez. "The World's Technological Capacity to Store, Communicate, and Compute Information." *Science* 332.6025 (2011): 60-65. ウェブサイト: <http://www.sciencemag.org/content/332/6025/60>
- ²³ Storage costs have now fallen from \$569 per gigabyte of storage in 1992 to \$0.02 per gigabyte in 2013 — at a rate of about 38 percent annually. Meeker, Mary. "Kleiner Perkins Internet Trends 2014." 2014. プレゼンテーション: <http://cryptome.org/2014/05/internet-trends-2014.pdf>
- ²⁴ Lohr, Steve. "For Big-Data Scientists, 'Janitor Work' Is Key Hurdle To Insights." *New York Times*. 2014: B4. 印刷物: http://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html?_r=0
- ²⁵ "Data, data everywhere," *The Economist* Feb. 25, 2010. <http://www.economist.com/node/15557443>
- ²⁶ Crovitz, L. Gordon. "Why 'Big Data' Is A Big Deal." *Wall Street Journal*. 2013. 印刷物: <http://online.wsj.com/news/articles/SB10001424127887324077704578364632408717740>
- ²⁷ Reports of droughts in Angola in 2006 triggered a warning about possible cholera outbreaks in the country, because previous events had taught the system that cholera outbreaks were more likely in years following droughts. The system's warnings were correct between 70 percent and 90 percent of the time. Simonite, Tom. "Software Predicts Tomorrow's News by Analyzing Today's And Yesterday's." *MIT Technology Review* 2013. 印刷物: <http://www.technologyreview.com/news/510191/software-predicts-tomorrows-news-by-analyzing-todays-and-yesterdays/>
- ²⁸ "Ten Big Data Case Studies in a Nutshell." *TechTarget*, 2013. 印刷物: 必須ガイド: <http://searchcio.techtarget.com/opinion/Ten-big-data-case-studies-in-a-nutshell>
- ²⁹ "Big Data to Predict Offshore Accidents, Tsunamis and Other Natural Disasters." *Predictive Analytics Today*. 2013 ウェブサイト: <http://www.predictiveanalyticstoday.com/big-data-predict-shore-accidents-tsunamis-natural-disasters/>
- ³⁰ "New Developments in Big Data Visualization." *USTelecom Media* 2014. ウェブサイト: <http://www.ustelecom.org/blog/new-developments-big-data-visualization#sthash.HefD5H52.dpuf>
- ³¹ Quinn, Tom. "New and Unexpected Uses for Scoring Technology." *Credit Score Blog* 2011. ウェブサイト: <http://blog.credit.com/2011/06/new-and-unexpected-uses-for-scoring-technology/>
- ³² Crovitz, L. Gordon. "Why 'Big Data' Is A Big Deal." *Wall Street Journal* 2013: 印刷物: <http://online.wsj.com/news/articles/SB10001424127887324077704578364632408717740>
- ³³ Data overload: Today's experienced clinician needs close to 2 million pieces of information to practice medicine and doctors subscribe to an average of seven journals, representing over 2,500 new articles each year, making it almost impossible to keep abreast with the latest information about diagnosis, prognosis, therapy and related health issues. "Clinical Decisions Support Systems: The Time Has Come..." Frost & Sullivan, 2009. 印刷物: マーケットインサイト: <http://www.frost.com/prod/servlet/cio/181298788>
- ³⁴ "Clinical Decisions Support Systems: The Time Has Come..." Frost & Sullivan, 2009. 印刷物: マーケットインサイト: <http://www.frost.com/prod/servlet/cio/181298788>
- ³⁵ "Kinect Sign Language Translator Expands Communication Possibilities." *Microsoft Research* 2013. ウェブサイト: <http://research.microsoft.com/en-us/collaboration/stories/kinect-sign-language-translator.aspx>

- ³⁶ “Brats 2012 - Multimodal Brain Tumor Segmentation Challenge.” *CodaLab*, 2012. 印刷物 : <https://www.codalab.org/competitions/191>
- ³⁷ Grady, Denise. “3-D Mammography Test Appears To Improve Breast Cancer Detection Rate.” *New York Times* 2014: p. A1 印刷物 : http://www.nytimes.com/2014/06/25/health/breast-cancer-3d-mammography-test-x-ray.html?emc=edit_th_20140625&nl=todaysheadlines&nlid=435891&r=0
- ³⁸ “The game-changing technology that’s transforming manufacturing.” *Manufacturing Weekly*, January 31, 2014. <http://web.archive.org/web/20140131233544/http://www.manufacturingweekly.com/supercomputers/>
- ³⁹ “The Return on the Data Asset in the Era of Big Data: Capturing the \$1.6 Trillion Data Dividend.” *Cloud Platform News Bytes Blog* 2015. ウェブサイト : <http://blogs.technet.com/b/stbnewsbytes/archive/2014/04/15/the-return-on-the-data-asset-in-the-era-of-big-data-capturing-the-1-6-trillion-data-dividend.aspx>
- ⁴⁰ Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making*. Cap Gemini, 2012. ウェブサイト : POV : <http://bigdata.pervasive.com/Solutions/Telecom-Analytics.aspx>
- ⁴¹ A 1 percent productivity increase may seem small, but as Jeff Immelt, CEO of GE puts it, “tell an oil guy you can use software to save him one percent on something, and that guy will be your friend for life.” Evans, Peter C., and Marco Annunziata. *Pushing the Boundaries of Minds and Machines*. GE, 2012. ウェブサイト <http://files.gereports.com/wp-content/uploads/2012/11/ge-industrial-internet-vision-paper.pdf>
- ⁴² Gartner, “Gartner Says Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally To Support Big Data By 2015.” 2012. 印刷物 : <http://www.gartner.com/newsroom/id/2207915>
- ⁴³ BSA/IPSOS Global Data Analytics Poll, November 2014, www.bsa.org/datasurvey
- ⁴⁴ BSA/IPSOS Global Data Analytics Poll, November 2014, www.bsa.org/datasurvey
- ⁴⁵ BSA/IPSOS Global Data Analytics Poll, November 2014, www.bsa.org/datasurvey
- ⁴⁶ According to ESG research, data managed per hospital is expected to increase from 168 terabytes in 2010 to 6 terabytes by 2015. “Digital Imaging in the Cloud.” *There Magazine* 2012: 16. 印刷物 : http://www.agfahealthcare.com/he/global/en/binaries/THERE_12_tcm541-95647.pdf
- ⁴⁷ Manyika, James et al. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 2011. 印刷物 : http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
- ⁴⁸ Researchers trained a machine-learning algorithm on data from 133,000 patients. The model still needs more work to reduce false positives. Rutkin, Aviva. “Machine Predicts Heart Attacks 4 Hours Before Doctors - New Scientist.” *New Scientist*. 2014. ウェブサイト : <http://www.newscientist.com/article/mg22329814.400-machine-predicts-heart-attacks-4-hours-before-doctors.html>
- ⁴⁹ Ford’s modern hybrid Fusion model generates up to 25 GB of data per hour. Hemsoth, Nicole. “How Ford Is Putting Hadoop Pedal To The Metal.” *Datanami*. 2013. ウェブサイト : http://www.datanami.com/2013/03/16/how_ford_is_putting_hadoop_pedal_to_the_metal/
- The Chevy Volt contains over 10 million lines of software code, and software developer is one of the fastest growing technical professions in Southeast Michigan, a region long known for its manufacturing prowess. Trop, Jaclyn. “Detroit, Embracing New Auto Technologies, Seeks App Builders.” *The New York Times*. June 30, 2013. <http://www.nytimes.com/2013/07/01/technology/detroit-embracing-new-auto-technologies-seeks-app-builders.html>
- ⁵⁰ Miller, Claire Cain. “If Robots Drove, How Much Safer Would Roads Be?” *New York Times* 2014: A3. 印刷物 : <http://www.nytimes.com/2014/06/10/upshot/if-robots-drove-how-much-safer-would-roads-be.html?ref=technology&r=0>
- ⁵¹ The 787 uses data sensors to reduce fuel, monitor systems, and even employs accelerometers in the nose of the plane to counteract turbulence. If the sensors register a sudden drop, they immediately tell the wing flaps to adjust (in a matter of nanoseconds) and in so doing, what used to be a 9 feet drop in an older plane can be reduced to just 3 feet in the 787, making for a much smoother flight. Gosling, Kevin. “E-Enabled Capabilities of the 787 Dreamliner.” *Aero Quarterly* 2009: 22-24. http://www.boeing.com/commercial/aeromagazine/articles/qtr_01_09/pdfs/AERO_Q109_article05.pdf
- ⁵² Jet engine maker GE says the engine data allows it to figure out things like possible defects 2,000 times as fast as it could before. Hardy, Quentin. “What Cars Did for Today’s World, Data May Do for Tomorrow’s?” *New York Times* 2014: B7. 印刷物 : http://bits.blogs.nytimes.com/2014/08/10/g-e-creates-a-data-lake-for-new-industrial-ecosystem/?_php=true&_type=blogs&_php=true&_type=blogs&module=BlogPost-Title&version=Blog%20Main&contentCollection=Big%20Data&action=Click&pgtype=Blogs®ion=Body&r=1&
- ⁵³ Long, Jessica, and William Brindley. *The Role of Big Data and Analytics in the Developing World*.

- Accenture, 2013. 印刷物 : Accenture Development Partnerships Insights into the Role of Technology in Addressing Development Challenges. https://www.accenture.com/us-en/~media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf
- ⁵⁴ The manufacturing sector stored nearly 2 exabytes of new data in 2010 alone. Manyika, James et al. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 2011. 印刷物 : http://www.mckinsey.com/~media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx
- ⁵⁵ Manyika, James et al. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 2011. 印刷物 : http://www.mckinsey.com/~media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx
- ⁵⁶ “‘Data Smart’ Strategies for Customers Are Yielding ‘Early But Impressive Returns.’” *Microsoft Research the Fire Hose* 2014. ウェブサイト : <http://blogs.microsoft.com/firehose/2014/05/22/data-smart-strategies-for-customers-are-yielding-early-but-impressive-returns/>
- ⁵⁷ Somers, Dan. “Manufacturing 4.0 – From Industrialization to Data-Driven Product Lifecycle.” *Citizenek*. 2013. ウェブサイト : <http://citizenek.com/2013/11/05/manufacturing-4-0-industrialisation-data-driven-product-lifecycle/>
- ⁵⁸ McKinsey reports that by using these data driven design techniques, Toyota was able to eliminate 80 percent of defects prior to building the first physical prototype. Manyika, James et al. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 2011. 印刷物 : http://www.mckinsey.com/~media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx
- ⁵⁹ Findings of the New Intelligent Enterprise Study. IBM 2010 New Intelligent Enterprise Global Executive Study. 2010. 印刷物 :
- ⁶⁰ Geron, Tomio. “Cows in the Cloud: The Hot Startup Moving Farmers into the Cloud.” *Forbes* 2012. Web; Helmer, Jodi. “Get Ready For Robot Farmers.” Yahoo 2014. ウェブサイト <https://www.yahoo.com/tech/get-ready-for-robot-farmers-100613764059.html>
- ⁶¹ De-Identification of Personally Identifiable Information, National Institute of Science and Technology, DRAFT NISTIR 8053 (April 2015).
- ⁶² Cavoukian, Ph.D., Ann, and El Emam, Ph.D., Khaled, Dispelling the Myths Surrounding De-Identification: Anonymization Remains a Strong Tool for Protecting Privacy, Information and Privacy Commissioner of Ontario, (June 2011); Cavoukian, Ph.D., Ann, and Daniel Castro. *Big Data And Innovation, Setting The Record Straight: De-Identification Does Work*. ITIF, 2014. 印刷版 : <http://www2.itif.org/2014-big-data-deidentification.pdf>
- ⁶³ See for example, Microsoft’s add on protecting privacy as their priority https://www.youtube.com/watch?feature=player_embedded&v=bt51MWI1oY
- ⁶⁴ Apple, *Government Information Requests*, noting that the company has incorporated state-of-the-art encryption into its iPhone operating system so that “your personal data such as photos, messages (including attachments), email, contacts, call history, iTunes content, notes, and reminders is placed under the protection of your passcode,” at www.apple.com/privacy/government-information-requests/
- ⁶⁵ Hachman, Mark, “Microsoft’s updated privacy policy makes it clear it’s not selling ads against your words,” *PCWorld*, June 11, 2014, <http://www.pcworld.com/article/2362130/microsofts-updated-privacy-policy-makes-it-clear-its-not-selling-ads-against-your-words.html>
- Timberg, Craig. “Newest Androids Will Join iPhone In Offering Default Encryption, Blocking Police.” *Washington Post* 2014: 印刷版 : <http://www.washingtonpost.com/blogs/the-switch/wp/2014/09/18/newest-androids-will-join-iphones-in-offering-default-encryption-blocking-police/>
- ⁶⁶ Data broker Acxiom opens consumer-facing data website, offers opt-out <http://cir.ca/news/acxiom-gives-consumers-data-peekaboutthedata.com>
- ⁶⁷ BSA/IPSOS Global Data Analytics Poll, November 2014, www.bsa.org/datasurvey
- ⁶⁸ McKinsey Global Institute. *Internet Matters: The Net’s Sweeping Impact On Growth, Jobs, And Prosperity*. McKinsey & Co., 2011. 印刷物.
- ⁶⁹ Manyika, James et al. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 2011. 印刷物 : http://www.mckinsey.com/~media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx
- ⁷⁰ According to Salaries of Data Scientists, an April 2014 study from Burtch Works.
- ⁷¹ Bagley, Rebecca. “How The Cloud And Big Data Are Changing Small Business.” *Forbes* 2014. ウェブサイト : <http://www.forbes.com/sites/rebeccabagley/2014/07/15/how-the-cloud-and-big-data-are-changing-small-business/>

- ⁷² BSA/IPSOS Global Data Analytics Poll, November 2014, www.bsa.org/datasurvey
- ⁷³ Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making*. Cap Gemini, 2012. ウェブサイト : POV. <http://bigdata.pervasive.com/Solutions/Telecom-Analytics.aspx>
- ⁷⁴ Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making*. Cap Gemini, 2012. ウェブサイト : POV. <http://bigdata.pervasive.com/Solutions/Telecom-Analytics.aspx>
- ⁷⁵ Gerbis, Nicholas. "10 Correlations That Are Not Causations." *HowStuffWorks*. 2015. ウェブサイト : <http://science.howstuffworks.com/innovation/science-questions/10-correlations-that-are-not-causations.htm>
- ⁷⁶ Vesset, Dan, Henry D. Morris, and John F. Gantz. *Capturing the \$1.6 Trillion Data Dividend*. IDC, 2014. 印刷物 : IDC White Paper.
- ⁷⁷ Westerman, George, Didier Bonnet, and Andrew McAfee, "The Advantages of Digital Maturity." November 2012, MIT Sloan.
- ⁷⁸ "Meeting the Big Data Challenge: Don't Be Objective." *Forbes* 2013. ウェブサイト : <http://www.forbes.com/sites/darden/2013/02/01/meeting-the-big-data-challenge-dont-be-objective/>
- ⁷⁹ IDG Enterprise's 2014 Big Data research. IDG. *CEOs Call for Big Data and IT Continues To Lead Investment Decisions*. 2014. 印刷物 : <http://www.idgenterprise.com/press/ceos-call-for-big-data-and-it-continues-to-lead-investment-decisions>
- ⁸⁰ Miller, Claire Cain. "If Robots Drove, How Much Safer Would Roads Be?" *New York Times* 2014: A3. 印刷物 : http://www.nytimes.com/2014/06/10/upshot/if-robots-drove-how-much-safer-would-roads-be.html?ref=technology&_r=0
- ⁸¹ Clemens, Samuel. "7 Facts about Data Quality [Infographic]." *InsightSquared*. January 3, 2012. ウェブサイト : <http://www.insightsquared.com/2012/01/7-facts-about-data-quality-infographic/>
- ⁸² Economist Intelligence Unit. *Big Data Harnessing a Game-Changing Asset*. SAS, 2011. ウェブサイト : http://www.sas.com/resources/asset/SAS_BigData_final.pdf
- ⁸³ "The Return on the Data Asset in the Era of Big Data: Capturing the \$1.6 Trillion Data Dividend." *Cloud Platform News Bytes Blog* 2015. ウェブサイト : <http://blogs.technet.com/b/stbnewsbytes/archive/2014/04/15/the-return-on-the-data-asset-in-the-era-of-big-data-capturing-the-1-6-trillion-data-dividend.aspx>
- ⁸⁴ BSA/IPSOS Global Data Analytics Poll, November 2014, www.bsa.org/datasurvey
- ⁸⁵ IT investments in the entire US economy, including retail, through the high-growth 1990s added 1 percent to 2 percent to the compound annual growth rate of US productivity. Farrell, Diana et al. *How IT Enables Productivity Growth*. San Francisco: McKinsey Global Institute High Tech Practice, 2002. 印刷物 : http://www.mckinsey.com/insights/business_technology/how_it_enables_productivity_growth
- ⁸⁶ National Bureau of Economic Research. NBER Working Paper No. 18315: *Is U.S. Economic Growth Over? Faltering Innovation Confronts The Six Headwinds*. 2012. 印刷物 : <http://www.nber.org/papers/w18315>
- ⁸⁷ Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making*. Cap Gemini, 2012. ウェブサイト : POV. <http://bigdata.pervasive.com/Solutions/Telecom-Analytics.aspx>
- ⁸⁸ Gertner, Joey. "GE for Making the 'Internet of Things' Real." *Fast Company* 2014. ウェブサイト : <http://www.fastcompany.com/most-innovative-companies/2014/ge>
- ⁸⁹ Evans, Peter C., and Marco Annunziata. *Pushing the Boundaries of Minds and Machines*. GE, 2012. ウェブサイト <http://files.gereports.com/wp-content/uploads/2012/11/ge-industrial-internet-vision-paper.pdf>
- ⁹⁰ "City Of Barcelona Realizes Vision of Innovative City Governance with Cloud, Devices, and Apps." *Customers.microsoft.com*. 2014. ウェブサイト : <https://customers.microsoft.com/Pages/Home.aspx>
- ⁹¹ "Autodesk the Gallery Masdar Headquarters Positive Energy Building." *Autodesk.com*. 2015. ウェブサイト : <http://www.autodesk.com/gallery/exhibits/currently-on-display/adrian-smith-gordon-gill-architecture-masdar-headquarters>
- ⁹² Bunge, Jacob. "Big Data Comes To The Farm, Sowing Mistrust." *Wall Street Journal* 2014. 印刷物 : <http://online.wsj.com/news/articles/SB10001424052702304450904579369283869192124>
- "Supply Chain Management Solution for Fast Moving Consumer Goods & Food Industries - Farm to Fork Tech Mahindra." *Techmahindra*. 2015. ウェブサイト : http://www.techmahindra.com/en-US/www/solutions/Pages/Enterprises/retail_farm_fork.aspx
- ⁹³ Between 2013 and 2020 the division of the digital universe between mature and emerging markets (e.g., China) will switch — from 60 percent accounted for by mature markets to 60 percent of the data in the digital universe coming from emerging markets. EMC Digital Universe. *Executive Summary Data Growth, Business Opportunities, and the IT Imperatives*. IDC, 2014. 印刷物 : <http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>

- ⁹⁴ Long, Jessica, William Brindley. *The Role of Big Data and Analytics in the Developing World*. Accenture, 2013. 印刷物 : Accenture Development Partnerships Insights into the Role of Technology in Addressing Development Challenges. https://www.accenture.com/us-en/~media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf
- ⁹⁵ Long, Jessica, William Brindley. *The Role of Big Data and Analytics in the Developing World*. Accenture, 2013. 印刷物 : Accenture Development Partnerships Insights into the Role of Technology in Addressing Development Challenges. https://www.accenture.com/us-en/~media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf
- ⁹⁶ Levy, Stephen. "Bill Gates and President Bill Clinton on the NSA, Safe Sex, and American Exceptionalism." *Wired* 2013: 印刷物 : <http://www.wired.com/2013/11/bill-gates-bill-clinton-wired/2/>
- Chhachhar, Abdul Razaque, and Siti Zobidah Omar. "Use of Mobile Phone among Fishermen for Marketing and Weather Information." *Archives Des Sciences* 65.8 (2012):107-119. 印刷物 : http://www.academia.edu/4592505/Use_of_Mobile_Phone_among_Fishermen_for_Marketing_and_weather_information
- ⁹⁷ Neuman, William, and Ralph Blumenthal. "New to the Archaeologist's Tool Kit: The Drone." *New York Times* 2014. 印刷物 : http://mobile.nytimes.com/2014/08/14/arts/design/drones-are-used-to-patrol-endangered-archaeological-sites.html?_r=1&referrer
- ⁹⁸ "Forty Thousand Relics to Be Digitized In Five Years." *Thanhni News*. 2010. ウェブサイト : <http://www.thanhniennews.com/entertainment/forty-thousand-relics-to-be-digitized-in-five-years-22816.html>
- ⁹⁹ Long, Jessica, William Brindley. *The Role of Big Data and Analytics in the Developing World*. Accenture, 2013. 印刷物 : Accenture Development Partnerships Insights into the Role of Technology in Addressing Development Challenges. https://www.accenture.com/us-en/~media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf
- ¹⁰⁰ Future of Privacy Forum. *Big Data: A Tool for Fighting Discrimination and Empowering Groups*. Future of Privacy Forum and Anti-Defamation League, 2014. 印刷物 : <http://www.futureofprivacy.org/wp-content/uploads/Big-Data-A-Tool-for-Fighting-Discrimination-and-Empowering-Groups-Report1.pdf>
- ¹⁰¹ Wladawsky-Berger, Irving. "The Changing Nature of Globalization in Our Hyperconnected, Knowledge-Intensive Economy." *Wall Street Journal* 2014. 印刷物 : http://blogs.wsj.com/cio/2014/06/20/the-changing-nature-of-globalization-in-our-hyperconnected-knowledge-intensive-economy/?mod=wsj_ciohome_cioreport
- ¹⁰² For example, Argentina, Australia, Brazil, Canada, Chile, China, Colombia, Costa Rica, Greece, Hong Kong, India, Indonesia, Korea, Mexico, Peru, Russia, Switzerland and Vietnam have adopted or have proposed rules that prohibit or significantly restrict companies from transferring personal information out of their respective domestic territories.

BSA | THE SOFTWARE ALLIANCE について

BSA | The Software Alliance (BSA | ザ・ソフトウェア・アライアンス: www.bsa.org) は、グローバル市場において世界のソフトウェア産業を牽引する業界団体です。BSA の加盟企業は世界中で最もイノベーティブな企業を中心に構成されており、経済の活性化とより良い現代社会を築くためのソフトウェア・ソリューションを創造しています。

ワシントン DC に本部を構え、世界 60 カ国以上で活動する BSA は、正規ソフトウェアの使用を促進するコンプライアンスプログラムの開発、技術革新の発展とデジタル経済の成長を推進する公共政策の支援に取り組んでいます。



www.bsa.org

BSA Worldwide Headquarters

20 F Street, NW
Suite 800
Washington, DC 20001

T: +1.202.872.5500
F: +1.202.872.5501

BSA Asia-Pacific

300 Beach Road
#25-08 The Concourse
Singapore 199555

T: +65.6292.2072
F: +65.6292.6369

BSA Europe, Middle East & Africa

2 Queen Anne's Gate Buildings
Dartmouth Street
London, SW1H 9BP
United Kingdom

T: +44.207.340.6080
F: +44.207.340.6090