原著論文

ビッグデータ（Big Data）の利活用による戦略的企業経営管理 ーその概念、現状、そして活用の経済的分析ー

成　耆政

Strategic Corporate Management by Use of Big Data

SUNG Kijung

要　旨

インターネットが日常化された最近の10余年の間、我々はデータ洪水（Data Deluge）の現象、情報の爆発時代に直面している。このような背景により生まれたビッグデータ（Big Data）という用語と概念は単に情報通信の技術分野のみならず、国や自治体はもちろん、医療、福祉、環境、観光、農業などさまざまな分野で活用が期待され、いろいろな意味合いで用いられている。とくに、最近にはビジネス分野で注目を浴びているが、これは、ビッグデータがグローバル経営環境下での企業活動、すなわちマーケティングや商品開発、業務改善など企業の経営戦略に甚大な影響を及ぼすからである。

キーワード

ビッグデータ（Big Data）　　経済的価値　　活用技術

目　次

Ⅰ．はじめに

1．問題提起

インターネットが日常化された最近の10余年の間、我々はデータ洪水（Data Deluge）[注1]の現象、情報の爆発時代に直面している。一般に、インターネットとウェブという技術が知られておおよそ20年が過ぎた今、我々は情報の洪水[注2]の中で暮らしている。日本でも2000年代に入り、ツイッター（Twitter）やフェイスブック（Facebook）などのSNSが急激に広がり、これにより「ビッグデータ（Big Data）」が社会全般にコアキーワードとして登場している[注3]。すなわち、情報処理の新しいパラダイムとして登場したビッグデータは未来の競争力を左右するコア概念でありつつある。

しかし、上述したように、このビッグデータという概念は新しいものではなく[注4]、1990年以降のインターネットの拡散より情報の洪水や情報爆発という概念として議論され、最近のビッグデータの概念として受け継いだものである。ビッグデータ（Big Data）[注5]という用語と概念は単に情報通信の技術分野のみならず、国や自治体はもちろん、医療、福祉、環境、観光、農業などさまざまな分野で活用が期待され、いろいろな意味合いで用いられている。とくに、最近にはビジネス分野で注目（図1）を浴びているが、これは、ビッグデータがグローバル経営環境下での企業活動、すなわちマーケティングや商品開発、業務改善、意思決定など企業の経営戦略に甚大な影響を及ぼすからである。

以上のことをふまえ、本稿では主に、ビッグデータの登場背景と概念的考察、市場状況や活用技術、そしてビッグデータ活用の事例分析をつうじて今後、企業や組織においてビッグデータを利活用する際に直面する課題などを考察する。

2．ビッグデータの登場背景

ここではビッグデータの登場背景について簡略に述べる。世界的なコンサルティング企業であるMcKinsey&Company[注6]（2011）はビッグデータの登場背景を次のようにまとめている。まず第1に、企業の顧客データトレッキングおよび収集行為の増加をあげている。顧客データがインターネットやスマートフォンなどの多様なメディアをつうじてトレッキングされ、オンラインのみならず、オフライン上でもユーザー情報、消費者行動に関する情報などの収集が可能になった。

第2に、貯蔵メディアとカメラモジュール、ディスプレー価格の引き下げなどはマルチメディアコンテン

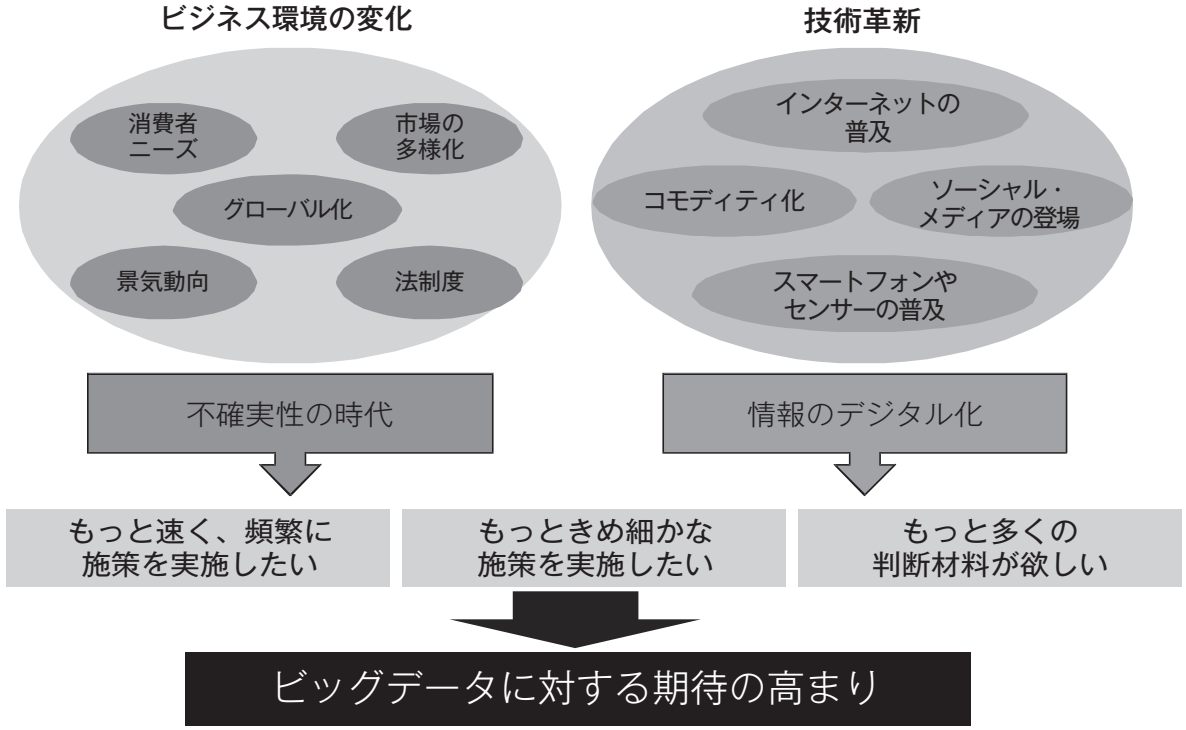


図1．ビッグデータが注目される理由

出所：ITR

ツの使用拡散と、これに関する情報の増加をもたらした。高画質の動画はすでにインターネットの全体トレピックの70%以上を占めているといわれ、今後、増え続けるであろう。

第3に、TwitterやFacebookなどSNSの急激な拡散とともに、テキストなどの非定型化データ[注7]の急増である。すなわち、Facebookでは、1カ月に1人のユーザーが平均90以上のコンテンツをアップロードし、YouTubeでは1分毎に24時間分量の動画がアップロードされている。

第4に、M2M[注8]やIoT（Internet of Things）[注9]などの通信技術の発展により通信ネットワークで発生するデータ量の増加である。すなわち、M2MやIoTなどの活性化をつうじてデータをユーザーが生成せず、インフラ自体が大量のデータを生成させるようになった。

また、城田（2012）によると、「なぜ今ビッグデータなのか？」とし、まず第1に、ビッグデータの民主化をあげている。これは、今のビッグデータは我々の日々の生活に密着した環境から生成されたことを意味している。

第2に、ハードウェアの価格性能比の向上、ソフトウェア技術の進化をあげている。すなわち、コンピュータの価格性能比の向上やハードディスク価格の下落、大量データを汎用品のサーバーで高速に処理できる「ハドゥープ（Hadoop）」[注10]の登場、さらにはクラウドコンピューティングの台頭などによってビッグデータを蓄積・処理しやすくなったからである。

第3に、クラウド（コンピューティング）[注11]の普及をあげている。すなわち、高性能のコンピュータやハードディスクなどのストレージ、そしてデータ分析のためのツールなどを自前で用意する必要性がなくなったことである。たとえば、Amazonのクラウドコンピューティングサービスである EC2[注12]やS3[注13]を使えば、大容量のデータ処理環境を構築しなくても分析が可能である。

以上をふまえ、ビッグデータの登場背景をまとめてみると、第1に、技術進歩によるデータの貯蔵・処理・分析能力の急激な拡大、第2に、データの貯蔵・処理費用の急激な減少、第3に、非構造化データの急激な増加と処理の必要性の向上、そして第4に、データの貯蔵方法のデジタル方式への転換の加速化などをあげることができる。すなわち、「インターネットの発展によりクラウド化が進み、その次の段階としてビッグデータに幅広い分野での応用が可能になった」[注14]ということである。

Ⅱ．ビッグデータの概念的考察

1．ビッグデータの意義

最近、多様なメディアでビッグデータに関する記事や論文などを見かけることが多くなってきている。

ビッグデータ[注15]について、明確に合意された定義はないものの、簡単にいえば、巨大なデジタルデータの総称であることは確かである。しかしながら、ビッグデータは単に巨大なデータのみを指すものではない。McKinsey（2011）によると、「ビッグデータとは、通常のデータベース管理ツールが貯蔵・管理および分析可能な範囲を超える規模のデータ」と定義づけている。すなわち、ビッグデータを既存のシステム、サービス、企業などで与えられた費用や時間で処理・分析できるデータの範囲を超える規模のデータのことといえる。

IT業界における市場調査およびコンサルティング企業であるIDC（International Data Corporation）は、次のいずれかの条件を満たすデータをビッグデータと定義している。すなわち、第1に、100TB以上のデータを有していること、第2に、音声や映像、金融取引情報、センサーなどのハイスピードストリーミングデータを利用していること、第3に、年率60%以上の成長率で生成されるデータであること、などである。そして、データを解析する際、スケーラブルなインフラを使用することも条件としている。そして、IDCはデータベースではなく、企業や組織の業務遂行に焦点を当てて、多様な種類の大規模データから安いコストで

表1．ビッグデータの定義（Ⅰ）

機　　関	ビッグデータの定義
Gartner（2011）	一般的に使われているハードウェア環境とソフトウェア・ツールでは、ユーザー層が許容できる時間内にキャプチャ・管理・処理できないデータ
McKinsey（2011）	典型的なデータベースソフトウェアのキャプチャ、格納、管理、分析能力を超えるサイズを持ったデータセットのこと
IDC（2011）	多様な種類の大規模データから安いコストで価値を抽出し、データの超高速収集・発掘・分析をサポートできるように考案された次世代技術およびアーキテクチャー

価値を抽出し、データの超高速収集・発掘・分析をサポートできるように考案された次世代技術およびアーキテクチャー[注16]と定義づけた（表1）。

ビッグデータについて、量的な側面と質的な側面[注17]に分けて述べることもできる。まず量的側面からみると、日々の生活において生成・処理される膨大なリアルタイム性のあるデータで、その容量は数テラバイト（TB、10^{12}バイト）から数ペタバイト（PB、10^{15}バイト）、数ゼタバイト（ZB、10^{21}バイト）までにのぼる。ただし、ビッグデータを量的側面からのみにアプローチすることは、あまり意味がない。

次に、質的側面からみると、ネットワーク接続端末の多様化などの技術の進歩により、ウェブ上ではさまざまなデータが処理されるようになった。SNSのテキストデータ、画像、音声、動画、位置情報、ログ情報（購入履歴や会員情報など）などのデータの種類の多様化、リアルタイム、ストリームなどのデータの生成頻度など、これらの多種多様、かつ大規模なデータがビッグデータになる（図2）。

2012（平成24）年版『情報通信白書』によると、ビッグデータについて次のように解説・定義されている[注18]。「ビッグデータとは何か。これについては、ビッグデータを「事業に役立つ知見を導出するためのデータ」とし、ビッグデータビジネスについて、「ビッグデータを用いて社会・経済の問題解決や、業務の付加価値向上を行う、あるいは支援する事業」と目的的に定義している例がある。ビッグデータは、どの程度のデータ規模かという量的側面だけでなく、どのようなデータから構成されるか、あるいはそのデータがどのように利用されるかという質的側面において、従来のシステムとは違いがあると考えられる」。

城田（2012）はビッグデータの定義について、狭義と広義のビッグデータに分けて次のように述べている。まず狭義のビッグデータとは、「既存の一般的な技術では管理するのが困難な大量のデータ群」としている。ここでの「既存の一般的な技術では管理するのが困難」とは、現在の企業データベースの主流を占めるリレーショナル・データベース（RD）では管理できない複雑な構造のデータを指したり、ボリュームが、増大したデータに対する問い合わせの応答時間が許容範囲を超えるような状態を招く膨大なデータを指すとしている。

次に、広義のビッグデータとは、「3V（Volume、Variety、Velocity）の面で管理が困難なデータおよびそれらを蓄積・処理・分析するための技術、それらのデータを分析し、有用な意味や洞察を引き出せる人材や組織を含む包括的な概念」と定義づけている。

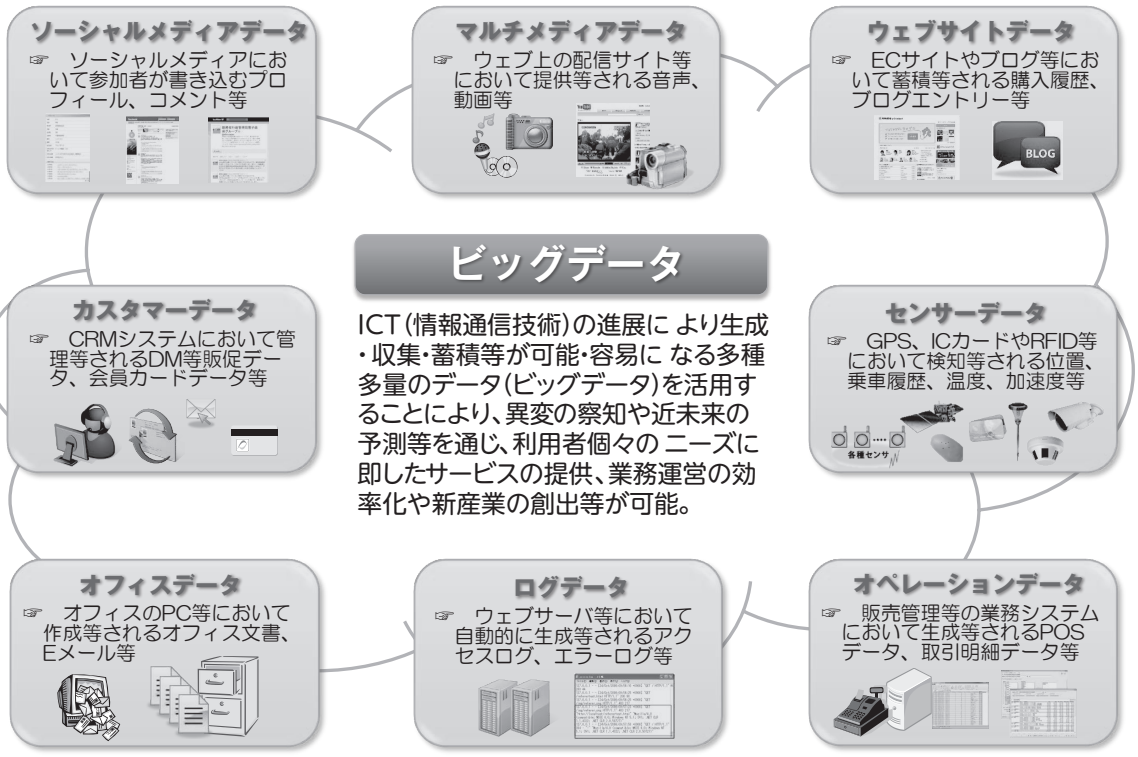


図2．ビッグデータを構成する各種データ（例）

出所：情報通信審議会ICT基本戦略ボード．ビッグデータの活用に関するアドホックグループ資料。

野村総合研究所（NRI）はビッグデータについて、広義的に第1に、人材･組織、第2に、データ処理・蓄積･分析技術、第3に、データなどの3要素として定義づけ、ビッグデータの特性である3Vは上記3要素の中でデータに当てはまる特性として狭義の定義として区分した[注19]（図3）。

網野（2013）はビッグデータの定義について、次の4つに大きく分けられるとしている（図4）。すなわち、第1に、単純にデータ量が大きいと述べているもの、第2に、データの種類を述べているもの、第3に、データの特徴を述べているもの、そして第4に、ビッグデータを使った分析全般の取り組みを概念としてビッグデータと呼んでいるものなどである。

Oracle[注20]によると、ビッグデータとは次の3つのタイプのデータを指している。まず第1に、従来のエンタープライズ･データである。これにはCRMシステム[注21]からの顧客情報、トランザクショナルERPデータ、Webストアのトランザクション、総勘定元帳データなどがあげら

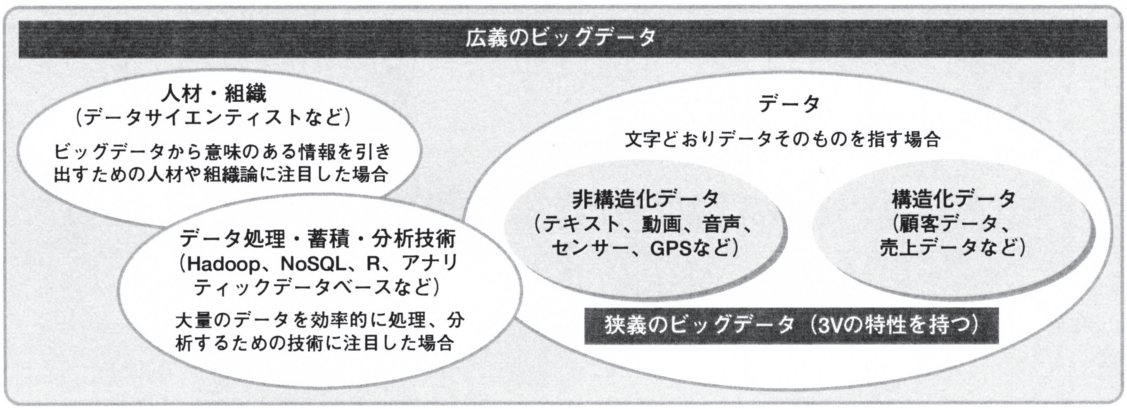


図3.　ビッグデータの定義（Ⅱ）

出所：NRI

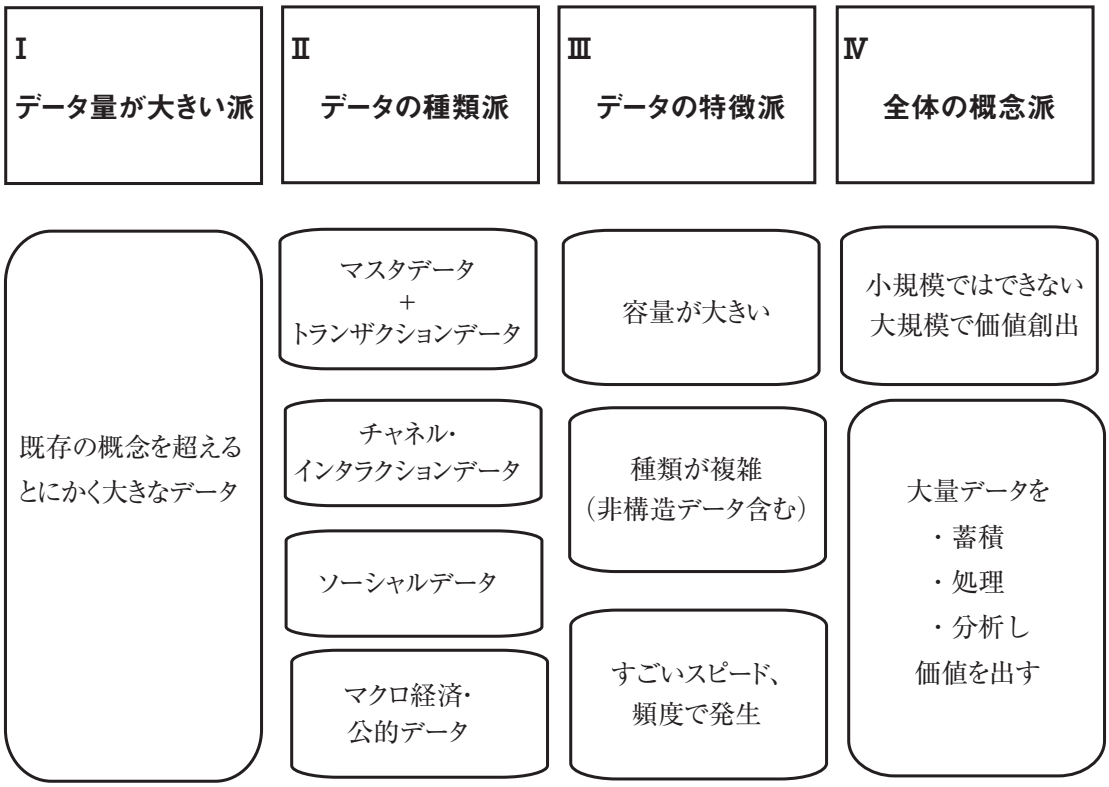


図4.　ビッグデータの定義（Ⅲ）

出所：網野（2013）、26頁。

れる。

第2に、機械が生成したデータ/センサー・データである。これにはCall Detail Record[注22]、Webログ、スマート・メーター、製造センサー、機器ログ、取引システム・データなどをあげることができる。

第3に、ソーシャル・データである。これには顧客フィードバック・ストリーム、Twitterなどのマイクロブログ・サイト、Facebookなどのソーシャル・メディア・プラットフォームなどをあげることができる。

2. ビッグデータの種類と構成

我々は日々さまざまなデータに囲まれて生活している。とくに、インターネットに代表されるネットワーク技術の急速な発展によりテキストや画像、動画などのデータが爆発的に増えている。

ビッグデータはデータの定型化（構造化）の程度により構造化データ（structured data）、半構造化データ（semi-structured data）、そして非構造化データ（unstructured data）などに分けることができる（表2）。

非構造化データはデータベースなどで管理しやすい構造化データに対するもので、申込書、契約書や報告書などの紙の文書、パソコンで作成されたオフィス文書、電子メールなどの通信文、音声、ウェブコンテンツ、音楽・写真・映像などのデジタル・コンテンツ、ファックス、スキャニングで得られた電子化文書などのデータのことをいう。このように整理しにくい非構造化データは、一般的には整理が簡単な構造化データの4倍以上はあるといわれている[注23]。電子メールやブログ、SNSなどでやりとりされる非構造化データは企業が抱えるデータの約80%を占めるといわれている[注24]（図5）。また、IBM報告書によると、全世界の情報の80%は非構造化データで、非構造化データの増加率は構造化

表2. ビッグデータの種類

区　分	内　　容
・構造化データ	関係型データベースやスプレッドシートなどのような固定されたフィールドに貯蔵されたデータ
・半構造化データ	XMLやHTMLテキストのように、固定されたフィールドに貯蔵されてないが、メタデータスキマーを含むデータ
・非構造化データ	テキスト分析が可能なテキスト文書やイメージ、動画、音声データなどの固定されたフィールドに貯蔵されてないデータ

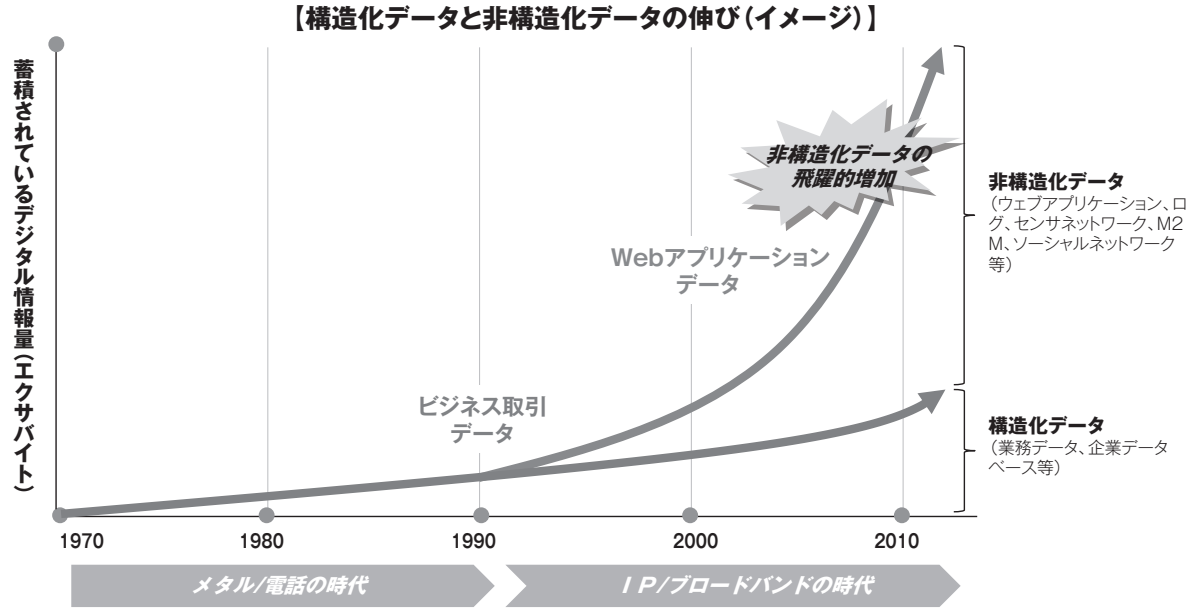


図5. 増大する非構造化データ

出所：総務省情報通信国際戦略局情報通信経済室.情報流通・蓄積量の計測手法に係る調査研究報告書（2013）、p.4。

データ増加率の15倍になるとしている（Paul Zikopoulos et al., 2012)。そして、2.5エクサバイト（exabyte：EB；10^{18} ＝ 百京バイト)もの大量データが日々生成されており、現代に存在するデータの90%はこの2年以内に生成されているとしている[注25]。

3. ビッグデータの特性

ここではまず既存のデータとビッグデータの相違点について簡略に述べていきたい（表3）。ビッグデータは既存のデータと比べ、第1に、10倍以上の多いデータを、第2に、ログデータ、購買記録など構造化データのみならず、ソーシャルメディア、位置情報、センサーなど非構造化データまで分析対象に含み、第3に、多様なデータ間の関係を同時に、可能な限り早く処理できるコンピューティング技術を適用し、第4に、多様で信頼できる分析結果を提示し価値を創出するデータ処理･分析方式である[注26]。

ビッグデータの特徴として、一般的に3V（Volume、Variety、Velocity）に1V（Veracity）や1C（Complexity）を追加して述べることができる（表4、図6）。

1) ビッグデータの規模・容量（Volume）：大量のデータを蓄積・処理可能

これは、収集され処理･分析されるデータ量が物理的に極めて大きいことを意味し、データの大きさのみならず、データが持つ属性や価値までも含む。スマートフォンやタブレット、SNS、M2Mの急激な普及に伴い、データの大きさは想像を絶するほど急増した。

2) ビッグデータの多様性・種類（Variety）：多様なデータに対応可能

企業や組織が保有するデータの中で、統一された構造として整理しにくい非構造化データの割合が90%を占め、テキスト、電子メール、写真、イメージ、動画、音声、株式データ、検索ファイル、コールセンター通話記録、センサー、ネットワークなど、多様な形態のデータを含むことを意味する。

3) ビッグデータの頻度・速度（Velocity）：高速・リアルタイムのデータ処理・転送可能

上記のビッグデータの多様性で述べたように、ビッグデータの大部分はウェブ検索ログ、センサーなどから持続的で、速いスピードで生成される。これはデータの生成と処理･分析のスピードを意味し、データの生成後、貯蔵されるまでの速度と、発生したデータの無意味な部分を処理する速度、そして生成されたデータを分析し意味を抽出するまでの速度を意味する。

表3. 既存のデータとビッグデータの相違点

区　　分	既存データの分析	ビッグデータ
データの量	・テラバイト水準	・ペタバイト水準（最小100テラバイト以上） ・クリックストリーム(注)データの場合、顧客情報の収集および分析を長期間にわたって遂行すべきなので、既存の方法とは比較し、処理すべきデータの量は膨大である
データの類型	・構造化データ中心	・ソーシャルメディアデータ、ログファイル、クリックストリームデータ、コールセンターログ、通信CDRログなど非構造化データの割合が高い ・処理の複雑性を増やす要因
プロセスおよび技術	・プロセスおよび技術が相対的に単純 ・処理･分析過程が構造化 ・原因・結果究明が中心	・多様なソース、複雑なロジック処理、大容量データ処理などにより処理が複雑すぎて、分散処理技術が必要 ・よく定義されたデータモデル･相関関係･手続きなどがなく、新しく多様な処理方法の開発が必要 ・相関関係の究明が中心 ・Hadoop、R、NoSQLなど開放型ソフトウェア

注：クリックストリームとは、Webページの訪問者が渡り歩いた軌跡のことである。インターネット上のページ移動の多くはリンクをクリックすることで行われるため、そのページにどのような経路で辿り着いたかということが「流れ（stream）」と表現されている（IT用語事典BINARYのウェブサイト資料）。

出所：ベ･ドンミンほか（2013）、P.41。

4) ビッグデータの真実性・正確性（Veracity）：
データの価値の探索

これは真実の度合い、すなわちデータがどれだけ真実を表しているのかということを意味する。これはデータ自体の特徴というよりも、データの取り方、データの発信媒体の信頼性などの側面も大きいであろう。

たとえば、センサーの故障によるノイズ、デマ情

表4. ビッグデータの特徴

区　　分	主　な　特　徴
ビッグデータの 規模・容量（Volume）	・IT技術の発展と日常化が進行され、毎年デジタル情報量が急激に増加
ビッグデータの 多様性・種類（Variety）	・ログ記録、SNS、位置、現実データなどデータ種類の増加 ・テキスト以外のマルチメディアなど非構造化データ類型の多様化
ビッグデータの 頻度・速度（Velocity）	・センサー、モニタリング、ストリーミング情報などリアルタイム性情報の増加 ・リアルタイム性によるデータ生成、移動速度の増加 ・大規模データ処理および価値あるリアルタイム活用のためのデータ処理および分析速度が重要
ビッグデータの 真実性・正確性（Veracity）	・データの矛盾、あいまいさによる不確実性、近似値を積み重ねた不正確さなどを排除して、本当に信頼できるデータによる意思決定が重要(注)
ビッグデータの 複雑性（Complexity）	・構造化されてないデータ、データ貯蔵方式の差、重複性の問題など ・データ種類の拡大、外部データの活用で管理対象の増加 ・データ管理および処理の複雑性が進化され新たな技法の要求

注：網野（2013）、p.24。
出所：Gartner、SAS、網野（2013）。

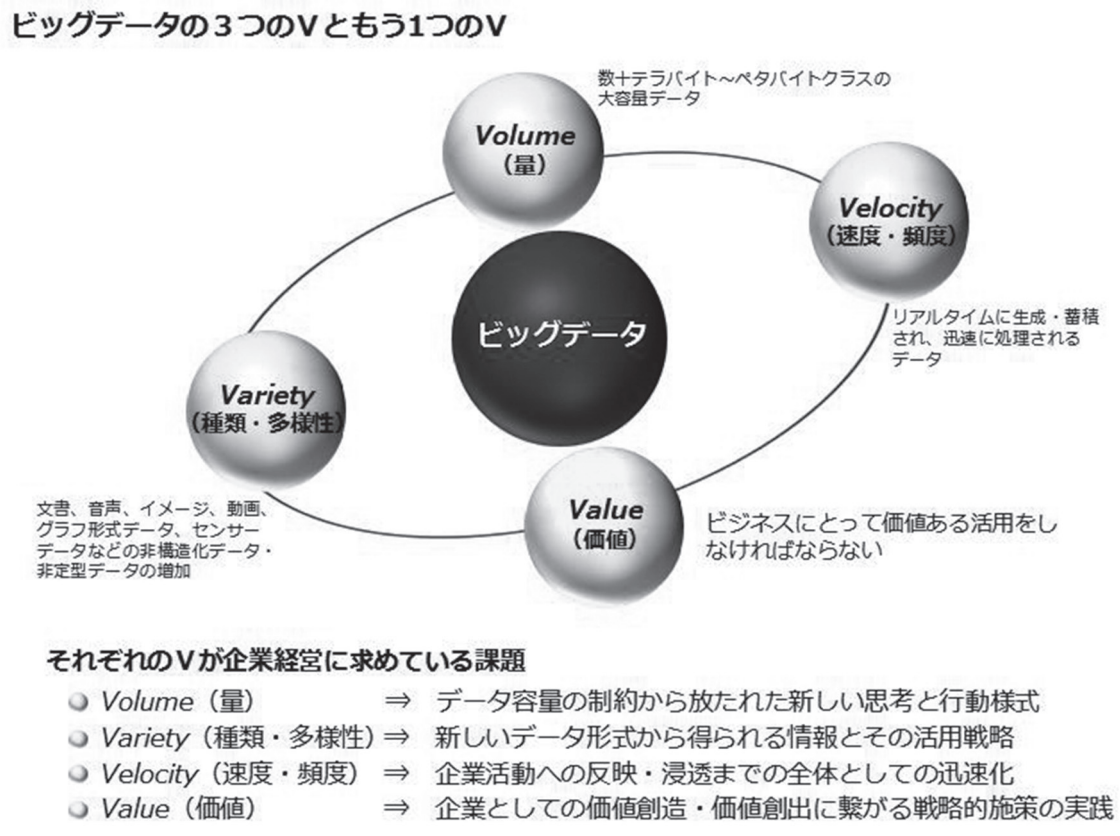


図6. ビッグデータの3Vと1V

出所：辻大志「企業経営から見たビッグデータの3つのV」ZDNet Japanのウェブサイト資料。

報の混じったブログやツイッターなどである。これはデータがどれだけ正しくて、どれだけ正しくないのかを把握した上で活用することが求められる重要な指標であろう[注27]。

5) ビッグデータの複雑性（Complexity）[注28]：データの複雑性への対応可能

ビッグデータの複雑性とは、データ構造やデータの獲得と処理にかかる速度、ドメインルール、貯蔵タイプなどデータの発生、処理などのプロセスを含むすべての要素が複雑になることを意味する。

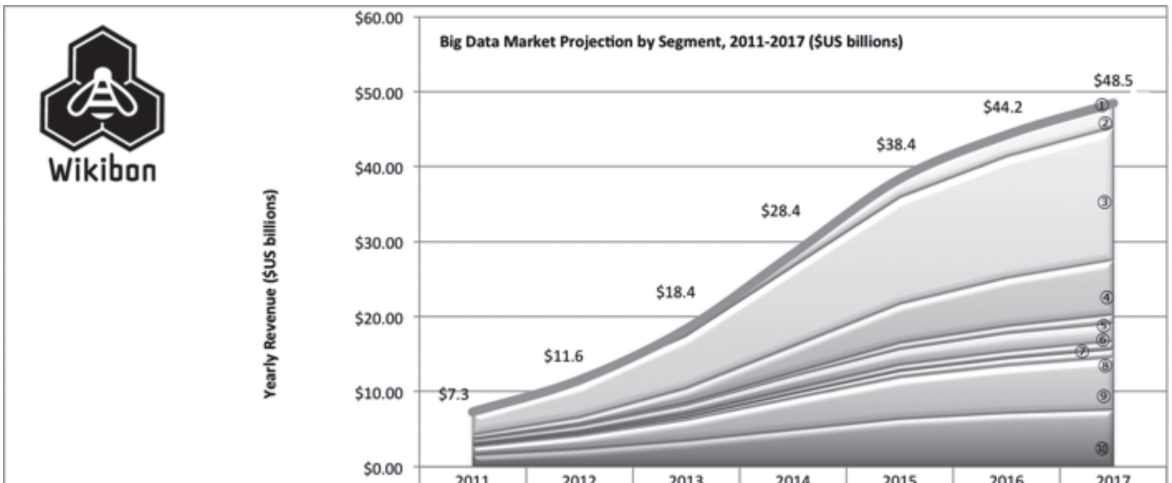
Ⅲ．ビッグデータの市場状況の考察 －ICT 市場に及ぼす経済的波及効果－

Gartnerは2012年と2013年に、ビッグデータを10大の戦略技術に選定した。また、他のグローバル調査機関もビッグデータ市場の成長を展望し、ビッグデータが全世界ICT市場に及ぼす経済的波及効果に注目し、ビッグデータが新たな情報社会のパラダイムを牽引することを期待している。

IDCは、全世界ビッグデータ市場が2010年の32億ドルから2013年に97億ドル、そして2015年には169億ドルで年平均39%の成長を予測し、2017年には324億ドル、2018年には415億ドルになると展望している（IDCのウェブサイト資料）。これは全体ICT市場の成長率の約6倍の値である。また、部門別成長率を見ると2015年にはソフトウェアとサービス部門がビッグデータ市場の約66%を占めると展望している。しかし、IDCはこのようなビッグデータに対する急激な需要にもかかわらず、企業はデータ分析に関する専門人材を確保できず、ビッグデータを分析・活用できない状況で、今後のビッグデータ市場の健全な発展に阻害要因として働くと予想している。この点についてはむすびの課題で少し詳しく述べることにする。

一方、ビッグデータの市場規模は市場の定義、範疇などにより調査・予測機関別に異なるが、注目すべきことはビッグデータ市場の成長率について高く展望していることである。

ITマネジメント担当者向けオンライン情報サイトの米Wikibon社が、「Big Data Vendor Revenue and Market Forecast 2011-2017」をリリースし、同データのインフォグラフィックを発表した（図7）。



		2011	2012	2013	2014	2015	2016	2017
Big Data XaaS Revenue	①	$0.36	$0.62	$1.07	$1.78	$2.52	$2.97	$3.31
Big Data Professional Services Revenue	②	$2.80	$4.42	$6.98	$10.62	$14.15	$16.17	$17.59
Big Data Application (Analytic and Transactional) Revenue	③	$0.52	$0.99	$1.90	$3.47	$5.29	$6.48	$7.38
Big Data NoSQL Database Revenue	④	$0.07	$0.13	$0.27	$0.50	$0.79	$0.98	$1.12
Big Data SQL Database Revenue	⑤	$0.62	$0.88	$1.25	$1.72	$2.14	$2.36	$2.51
Big Data Infrastructure Software Revenue	⑥	$0.14	$0.24	$0.40	$0.64	$0.88	$1.03	$1.14
Big Data Networking Revenue	⑦	$0.15	$0.23	$0.37	$0.56	$0.75	$0.86	$0.93
Big Data Storage Revenue	⑧	$1.10	$1.75	$2.76	$4.20	$5.59	$6.39	$6.95
Big Data Compute Revenue	⑨	$1.53	$2.29	$3.40	$4.89	$6.26	$7.01	$7.53
Total Big Data Revenue	⑩	$7.3	$11.6	$18.4	$28.4	$38.4	$44.2	$48.5

図7.　ビッグデータ市場予想

出所：米Wikibon社. Big Data Vendor Revenue and Market Forecast 2011-2017. 一部改ざん･引用。

ビッグデータ市場は2012年には114億ドルに到達し、2013年には181億ドル、2017年には470億ドルに到達すると予測し、2012年から2017年の5年間約31%の年間成長を達成すると予想している。

また、ビッグデータ市場をXaaS[注29]、専門家サービス（Professional Services）、応用ソフトウェア（Application（Analytic and Transactional）Software）、NoSQLデータベースソフトウェア、SQLデータベースソフトウェア、インフラストラクチャーソフトウェア（Infrastructure Software）、ネットワーキング（Networking）、ストリッジ（Storage）、そして計算（Compute）などに細分化し、各市場についても予想している。

同レポートでは、ビッグデータ市場はまだ初期導入段階で、大幅な成長が見込めるとしている。主要収入領域は今後5年間でビッグデータインフラストラクチャから付加価値サービスやソフトウェアへ変化し、ビッグデータインフラストラクチャ/ミドルウェア/技術サービスは徐々にコモディティ化すると予測している。

以上のように、ビッグデータ市場の構造上、各産業別の割合を探ってみると、市場調査機関別に多少の差はあるものの、ビッグデータサービス部門が41.5～44.0%、ハードウェア部門が28.9～31.0%、そしてソフトウェア部門が25.0～29.7%を占め、ビッグデータサービス部門が最も高い割合を占めていることが分かる。これは、ビッグデータの主な技術が具現・適用される、サービス領域が全体ビッグデータ市場で最も重要であることを指している。

ビッグデータが経済成長に寄与する可能性について、その寄与を具体的に分析するためには、その前提として、日本においてどの程度のビッグデータが生成・流通・蓄積されているのか、その実態を把握することが極めて重要なことである。ここでは特許庁の「平成25年度特許出願技術動向調査報告書（概要）－ビッグデータ分析技術－」の資料を用いて述べることにする。日本国内におけるビッグデータ市場の状況については、日本におけるビッグデータ分析市場の規模を測る指標となる2012年のデータ流通量のメディア別推移を＜図8＞に、2012年のデータ蓄積量を＜図9＞のように示す。

この資料によると、2012年のデータ流通量は約2.2エクサバイト（その内訳は、電子メール238PB、RFIDデータ584PB、GPSデータ348、固定IP電話178PB、そしてPOSデータ765PBなどである）、蓄積量は約9.7エクサバイト（その内訳は、商業3.2EB、サービス2.5EB、建設1.29EB、製造業0.826EB、金融・保険0.82EBなどの順である）であった。そしてデータ流通量は2005年から7年間で約5倍に増加し、データ蓄積量は同年のデータ流通量の約5倍となっており、蓄積量には過去からの累積が含まれることを考慮しても、蓄積量に対する流通量の比率は小さく、同一企業内などでのデータ活用が多く、データ流通が進んでいない状況に留まっていると考えられるとしている。また、＜図8＞に示したデータ流通量の推移を基に将来の伸びを予想した結果を＜図10＞のように示す。これによると、2008年から2012年までの4年間の伸び率は約1.21倍/年、

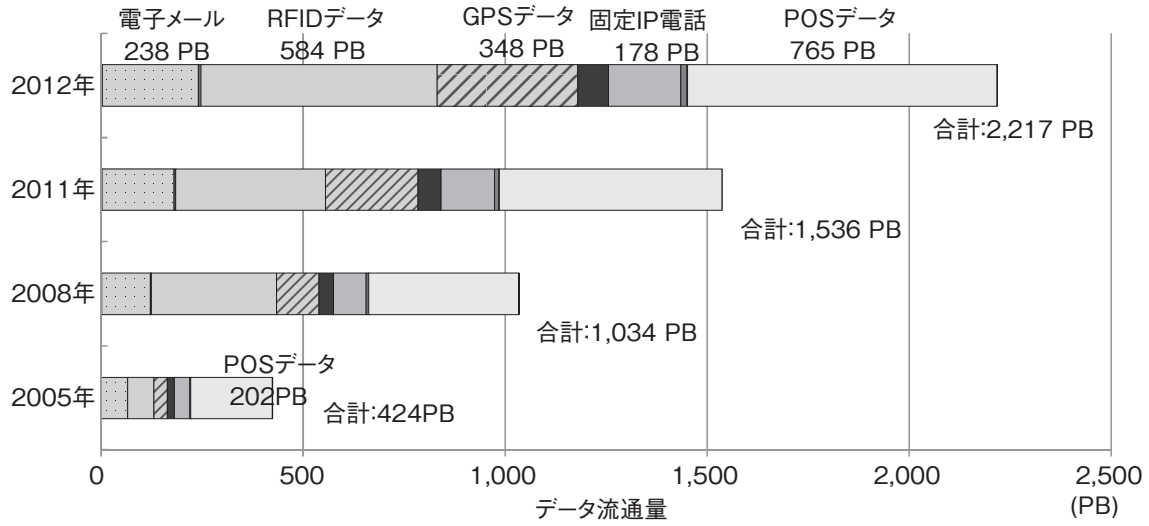


図8.　日本におけるデータの流通量

出所：特許庁. 平成25年度特許出願技術動向調査報告書（概要）－ビッグデータ分析技術－（2014）、P.7。

2011年から2012年の1年間では約1.44倍/年となっている。これらの伸び率が続くと仮定すると、2020年のデータ流通量は、最小で約10.3エクサバイト（2012年比4.7倍)、最大で約41.7エクサバイト（2012年比19.0倍)となると予想することができる。

Ⅳ．ビッグデータの活用技術

ビッグデータの活用に関する技術[注30]は日々進化をなし遂げている。ビッグデータ技術とは大量のデータを分析し、サービスへの付加価値を見つけるための技術である。したがって、ビッグデータを効率的に分析することができるようなシステム技術の重要性がますます大きくなっている。ビッグデータの特性をふまえ、ビッグデータを活用するためには新しい技法の分析手法を導入すべきである。最近、ビッグテーブル、カサンドラ、データウェアハウス、分散システム、グーグルファイルシステム、ハドゥープ、Hベース、マップリデュースなどのビッグデータ関連技術が開発されている（表5)。ここではハドゥープ、NoSQL、データウェアハウスについて少し詳しく述べることにする。

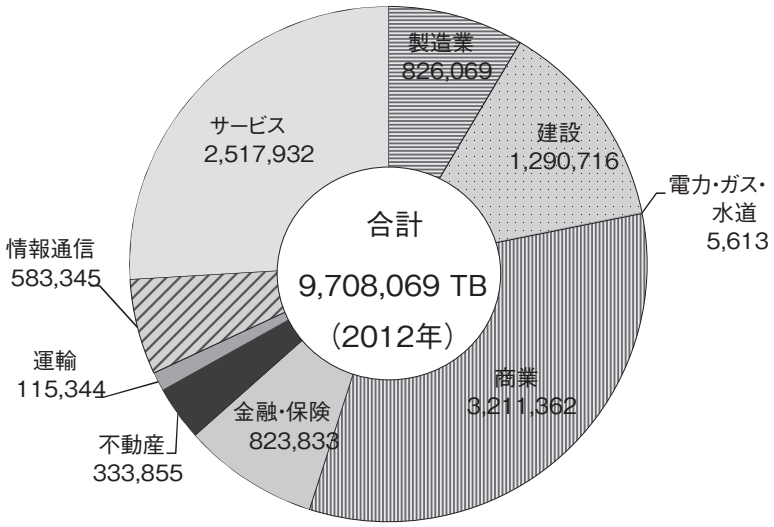


図9.　日本におけるデータの蓄積量(単位：TB)

出所：<図8>と同じ。

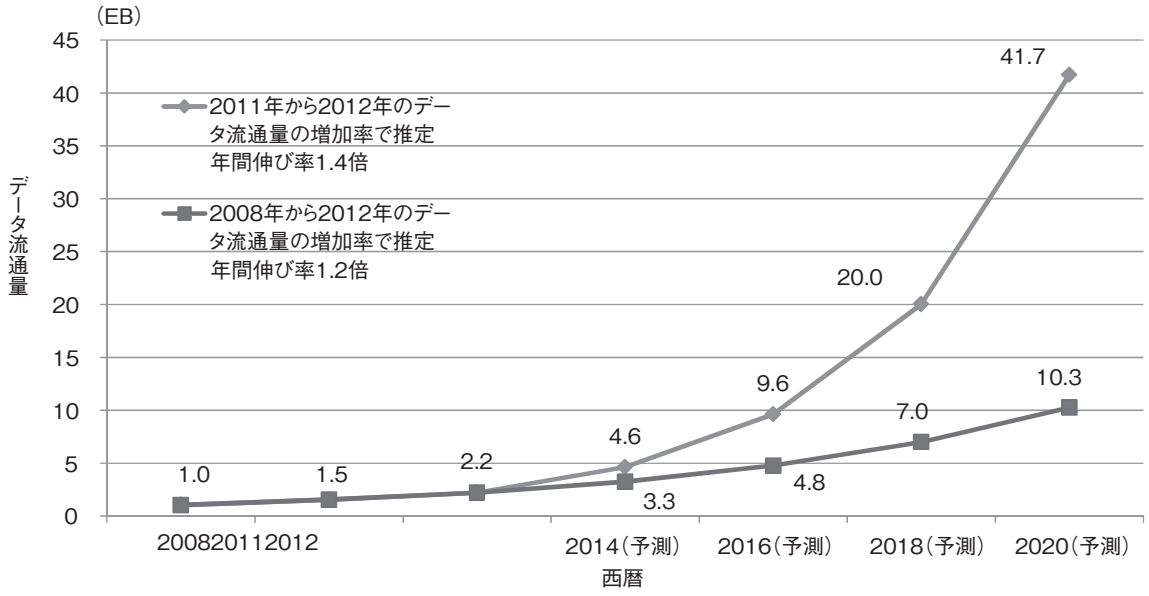


図10.　日本におけるデータの流通量の推移と予測

出所：<図8>と同じ資料のp.8。

1. 大規模データの効率的な分散処理フレームワークのハドゥープ（Hadoop）技術

ハドゥープ[注31]とは、アメリカのアパッチ（Apache）[注32]が大規模データの効率的な分散処理（複数のサーバーに分散して処理）などを支援するために開発したオープンソースソフトウェアツールセットとフレームワーク（open-source software tool set and framework）[注33]である。これは、検索システムGoogleのコンポーネントである分散ファイルシステムのGFS（Google File System）」、分散ロックシステムの「Chubby」、並列プログラミングモデルの「MapReduce」、キー・バリュー型データストアの「BigTable」、プログラミング言語の「Sawzall」などがオープンソース化されたものである[注34]。

そして、分散されたデータをハドゥープの上では統合された形で管理することができることで、データが大規模で、複雑になっても対応できるという高い拡張性を備えている[注35]。これはGoogleのGFS（Google File System）[注36]に基づき、ダグ・カッティング（Doug Cutting）により開発され、関連する他のオープンソースプロジェクトと、ハドゥープ生態系を形成し、ビッグデータの収集・貯蔵・処理の標準となった。

基本的にハドゥープフレームワークは数多くの汎用コンピュータを非共有方式で構築したクラスター上で、データをHDFS[注37]に分散貯蔵した後、マップリデュースというユーザー定義関数の実行をつうじてデータをバッチ方式で処理する。ハドゥープは大きく次の3つの構成要素となっている。第1に、外部の大規模データを集めてくる「コレクター」、第2に、集合されたデータをデータベースに「貯蔵する領域」、そして第3に、貯蔵されたデータに価値を付与

表5. ビッグデータのための技術の概要

技　術（Technology）	定　　義（Difinition）
・ハドゥープ（Hadoop）	複数の並行サーバーでビッグデータの処理が可能なオープンソースソフトウェア
・マップリデュース（MapReduce）	ハドゥープに依拠したアーキテクチャフレームワーク
・スクリプト言語（Scripting languages）	ビッグデータと相性が良いプログラミング言語（たとえば、Python[注1]、Pig[注2]、Hive[注3]）
・機械学習（Machine learning）	あるデータセットに最もフィットするモデルを高速で検索・発見するソフトウェア
・視覚的分析（Visual analytics）	視覚的、グラフィックフォーマットで分析結果を提示
・自然言語処理（Natural Language Processing;NLP）	テキスト分析（頻度や意味など）のためのソフトウェア
・インメモリ分析（In-memory analytics）	コンピュータメモリ上で、高速でビッグデータを分析

注1：Pythonは、Windows、Linux/Unix、Mac OS Xなどの主要なOSはもちろん、JavaやNETなどの仮想環境でも動作するプログラミング言語で、OSIに認証されたオープンソースライセンスで公開され、商用製品の開発にも無料で利用できる言語である。Pythonには、①クリーンで読みやすい文法、②強力なイントロスペクション機能、③直感的なオブジェクト指向、④手続き型のコードによる自然な表現、⑤パッケージの階層化もサポートした完全なモジュール化サポート、⑥例外ベースのエラーハンドリング、⑦高レベルな動的データ型、⑧事実上すべてのタスクをこなせる広範囲に及ぶ標準ライブラリとサードパーティのモジュール、⑨拡張とモジュールはC/C++で書くのが容易、⑩アプリケーションに組み込んでスクリプトインタフェースとして利用することが可能（Pythonのウェブサイト資料）などの特徴がある。

注2：PigはMapReduceのラッパーであり、簡単なデータフローを記述するだけでMap関数とReduce関数に変換し、Hadoop上で分散処理を実行することが可能なソフトウェアである。これは、2008年9月に米Yahoo社により公開されたオープンソースソフトウエアで、現在は、Hadoopサブプロジェクトとして開発が進められている。Pigは処理言語である「PigLatin」と、その実行環境で構成されている（IT proのウェブサイト資料）。

注3：Hiveの特徴は、MapReduceの処理をリレーショナルデータベース（RDB）のテーブル操作のように実行できることで、Hiveの問い合わせ言語である「HiveQL」は、RDBの「SQL」に似ている。Hiveで扱えるオペレータには抽出の「SELECT」、結合の「JOIN」、グループ化の「GROUP」、そして集約の「UNION」などがある（IT proのウェブサイト資料）。

出所：Davenport（2014）, p.114.

する「分析領域」などである。

Hadoopは上述したように、オープンソースソフトウェアなので、IBM、マイクロソフト、日立製作所、富士通、NTTデータなどの企業が関連製品やサービスを提供している[注38]。

2. 非関係型データ貯蔵技術のNoSQL

伝統的なRDBMSではクラウドコンピューティング環境で発生するビッグデータを効果的に貯蔵・管理するのにさまざまな問題が生じる。この問題点を補完するために開発されたのがNoSQL（Not only SQL、No SQL;ノー・エスキューエル）[注39]である。NoSQLは、既存のRDBMS形態のリレーショナルデータベース管理システムとは異なる設計による非関係型データ貯蔵技術を意味する。すなわち、NoSQLはRDBMS製品群であるMS-SQL、Oracle、Sybase、MySQLなどのように共通のデータ貯蔵方式（table）とアプローチ方法（SQL）を持つ製品群ではなく、RDBMSとは異なる形態のデータ貯蔵構造の総称である。しかし、製品によって各々その特性が異なるのでNoSQLを一つの製品群として定義することは難しい。

RDBMSが定型データの処理を必要とする業務システムでの利用に適しているのに対し、NoSQLはセンサーやソーシャルメディアなどの非定型データを含む多様なデータを大量にデータベース化するために利用されている[注40]。

NoSQLの特徴としては、第1に、クラウドコンピューティングに適している。第2に、柔軟なデータモデルである、そして第3に、ビッグデータの分析・処理に効果的である、などをあげることができる。

このNoSQLという技術で作られた製品の種類は150を超えるが、主なものとしてはOracleNoSQL、MongoDB、Casandra、BigTable、Hbase、Cough DB、Cloudataなどをあげることができる。

3. 戦略構築と意思決定を効率的支援するためのデータの貯蔵空間のデータウェアハウス

データウェアハウス（Data Warehouse: DWH）[注41]は1990年ウィリアム・インモン（William Inmon）により提唱され、ホストコンピュータが持っているデータを統合し、データを抽出・加工・要約し、経営活動に有用な情報を提供するための一連の情報処理技術である。すなわち、データウェアハウスとは企業や組織で一定期間情報システムを運用し、蓄積した基幹系業務データと外部データをサブジェクト（subject、主題）ごとに統合し、多様な分析を提供することでユーザーの戦略樹立や意思決定を効率的支援するためのデータの貯蔵空間（貯蔵庫）である。

データウェアハウスは単純にデータをかき集めただけのデータベースとは異なり、次のような特徴を備えている[注42]。まず第1に、サブジェクト指向性（subject oriented）があげられる。これは企業や組織の意思決定に必要な特定のサブジェクト、すなわち主要業務プロセス機能と関連した主題領域別にデータを構成することを意味する。たとえば、保険会社の場合、既存のプロセス中心のシステムでは自動車保険、生命保険、健康保険、傷害保険などに該当するが、これをサブジェクト領域別に見ると顧客、約款、請求、保険料などになる。

第2に、データウェアハウスの顕著な特徴として統合性（integrated）をあげることができる。既存のシステムは部署や部門、または組織別に一貫性のない大量のデータを重複管理するが、データウェアハウスは属性の名前、コードの構造、単位など一貫性を維持し、全社的観点で一つに統合された概念である。

第3の重要な特徴としては不揮発性（non-volatile；恒常性）があげられる。すなわち、データウェアハウスではデータを消さない、更新しないことを意味する。既存のデータベースでは追加、削除、変更などのような更新作業を持続的に実施するが、データウェアハウスは特別な場合を除いて、データを修正、削除せずに読み取り専用（read only）として保管する。

たとえば、各種分析を行う際には、履歴が大切な意味を持つこともあって、顧客の住所変更があった場合は、古いデータはそのまま残して新しい住所データを最新住所として追記する。また、間違った売上データを取り消す場合は間違ったデータを消去するのではなく、「間違った売上データを取り消す」という意味を持ったデータを追加するようにし、取消があったという事実を失わないようにするのが原則である[注43]。

第4に、データウェアハウスの特徴として時系列性（time variant）があげられる。データウェアハウスでは一定期間収集したデータを更新せずに貯蔵し、日、月、分期、年などのような期間関連情報を一緒に貯蔵する特徴がある。

Ⅴ．ビッグデータの活用とその事例分析

1．ビッグデータの経済的価値の展望

ビッグデータを活用可能な分野や範囲には、制限がないといえる。また、ビッグデータの活用においては、どのような経済的効果（表6）が得られるかを左右するので、その活用の目的を明確にしておくことが極めて重要である。そして、企業や組織においてはビッグデータを活用することで、コストの削減やデータ処理時間、意思決定時間の大幅な短縮や意思決定の質の大幅な向上、新製品や新サービスの開発、自社ブランド価値の向上などに大きな効果をもたらすことができる。

McKinseyによる[注44]と、ビッグデータをアメリカのヘルスケア部門、ヨーロッパの公共行政部門、アメリカの小売部門、グローバルな製造業部門、そしてグローバルな個人位置情報データ部門に適用際に1%の追加生産性の向上が可能で、各部門別に少なくとも1,000億ドルから7,000億ドル規模の経済的効果の創出を予想している。そして、生産性向上の程度により分けてみると、コンピュータ、電子製品および情報通信分野でビッグデータの適用効果が大きいと分析している。ビッグデータの経済的活用は産業部門別に約0.5から1%程度の生産性の増加をもたらす。たとえば、アメリカのヘルスケア部門では年間3,300億ドル、ヨーロッパの公共部門では2,500億ユーロを節減できるとしている。アメリカの小売業では、生産性は0.5%増加、売上純利益は60%以上増加、グローバルな製造業部門では開発コストが25%減少、製品の市場投入までの期間が20%から50%短縮、利益マージンが2%から3%増加、オペレーションコストが10%から25%削減、そして7%の収入増、グローバルな位置情報サービス部門では、2020年までに累計7,000億ドルから8,200億ドルの経済効果が創出されるなどの結果が出ている。

そして、ビッグデータの利活用による発現効果[注45]としては、まず第1に、ビッグデータの利活用による企業や組織の業務効率化と付加価値の創出をあげることができる。これについて、総務省の2012（平成24）年版『情報通信白書』（図11）によると、①医療部門での医療費最適化（3.1～4.6兆円）、②行政部門での行政効率化（7,200億円～1.2兆円）、社会保障給付是正（2,995.5億円～1.2兆円）、租税増収（2,133.9～8,535.6億円）、③小売部門での利益増加（0.95兆円以上）、④製造部門での製品開発費節減（最大5.7兆円）、⑤位置情報部門でのサービス収入（3,040億円）、そして⑥交通分野でのプローブ交通情報導入による渋滞解消効果（2.09兆円）をあげており、今後少なくとも10兆円規模の付加価値創出および12～15兆円規模の社会的コスト削減の効果があると考えられている。

第2に、パーソナル情報の市場創出効果をあげることができる。匿名パーソナル情報の市場規模は有望ビジネス分野として医療、安全、金融、運輸、小売、そしてサービスなどの分野に分けて、全体で約11,635億円規模と推計され、約4,905億円の金融分野と3,065億円の小売分野の市場規模が特に大きいとしている[注46]。金融分野における主なサービスとしては、複数の信用情報を統合し、信用リスクを分析、ヘッジすることと匿名化トレーディング情報活用サービスがあり、小売分野では仮想店舗の

表6．ビッグデータの活用価値

分　　野	主　な　内　容
・ヘルスケア部門（米）	・毎年3,000億ドルの価値 ・年～0.7%の生産性の増加
・公共行政部門（ヨーロッパ）	・毎年2,500億ユーロの価値 ・年～0.5%の生産性の増加
・小売業部門（米）	・利潤60%増加可能 ・年0.5～1%生産性の増加
・製造業	・製品開発費50%減少 ・運転資本7%節減可能
・個人位置データ（グローバル）	・サービス供給者の売上は1,000億ドル以上 ・エンドユーザーへの価値は7,000億ドル

出所：McKinsey Global Institute（2011）, p.12.

購入履歴活用サービスとリアル店舗の販売情報活用サービスなどがある。

第3に、ビッグデータの利活用に伴う新たなICT技術やソリューションの創出があげられる。上記の情報通信白書によると、データ収集でM2M（2020年に約9,000億円）、情報管理でクラウドサービス（2016年に2.8兆円、2020年に4.2兆円）とストレージ関連ソフトウェア（2020年に約977億円）、そしてデータ分析でビジネスインテリジェンスツール（2020年に約1,940億円）などのICT技術やソリューションを生み出すとしている。

そして、ビッグデータの社会･経済的意味とその経済的価値展望については各々、<表7>と<表8>のとおりである。

2. 日本のビッグデータ推進戦略

日本政府は2012年からビッグデータに関する政策を推進し、IT戦略本部をつうじて、新しい情報通信技術戦略を公表している。すなわち、クラウドコンピューティングサービスの競争力の確保の工程表ではビッグデータビジネス創出のためのM2M通信技術の開発と標準化など環境整備の実施している。また、2012年7月に公表した「日本再生に向けた改革工程表」の中で科学技術イノベーション・情報通信戦略では2015年まで実施すべき事項として、情報通信技術を活用した異分野融合により、新たに2兆円程度の市場創出を目標としている。そして、2020年までに情報通信技術を活用した異分野融合により、約9兆円規模の関連市場を創出することを実現すべき成果目標としている。

IT総合戦略本部では、IT･情報資源の利活用で、未来を創造する国家ビジョンとして、「世界最先端IT国家創造宣言」（平成25年6月14日閣議決定、平成26年6月24日改定）を策定した。この宣言では、2020年までに世界最高水準のIT利活用社会を実現することを目標に、第1に、革新的な新産業・新サービスの創出と全産業の成長を促進する社会の実現、第2に、国民が健康で安心して快適に生活できる世界一安全で災害に強い社会の実現、第3に、公共サービスがワンストップで誰でも、どこでも、いつでも、受けられることができる社会の実現などの3項目について目指すべき社会・姿を明らかにし、その実現に必要な取り組みなどをとりまとめている。

この宣言において示された目指すべき社会・姿の実現に向けて、どの府省が、いつまでに、具体的に何を実施するのかを明らかにするとともに、各府省間での連携が必要な施策については、個々の役割分担と達成すべき事項を明確化することにより、着実に具体的な成果に結び付けることを目的として工程表を策定してある。また、この宣言で示された取り組みや目標に対して、短期、中期、長期に分けて、各府省が実施する施策を明示している。ビッグデータの活用に関する推進戦略については上記の第1目標である「革新的な新産業・新サービスの創出と全産業の成長を促進する社会の実現」の実施事項として提示されている[注47]。

日本におけるビッグデータの活用を推進するための具体的方策については<表9>のとおりである。

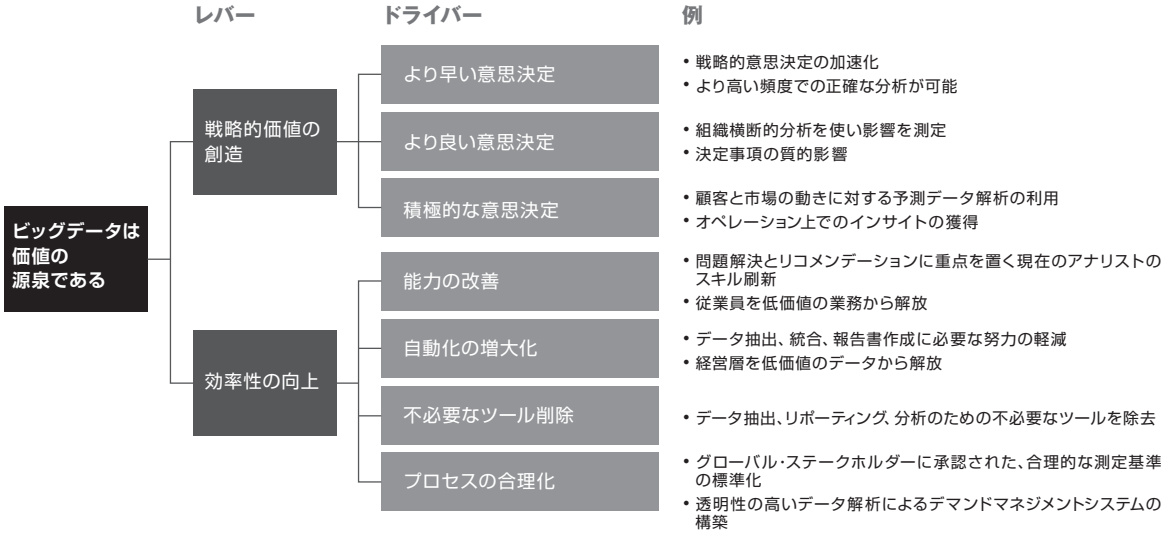


図11.　ビッグデータ活用の効果

出所：A.T.Kearney analysis

3. ビッグデータ活用の事例分析

1) Googleのビッグデータ戦略

周知のごとく、Googleはウェブサイト検索、クラウドコンピューティング、そして広告を主なビジネス領域とするアメリカの多国籍企業である。Googleの使命は世界中の情報を整理し、世界中の人々がアクセスできるようにすることである。

Googleは検索サービス[注48]によりユーザーの関心事項を収集している。たとえば、Googleは検索サイトをつうじてインターネット上のウェブページを、Gmail、Picasa（ピカサ）やカレンダーなどの無料サービスをつうじてユーザーのデータを、ストリートビューやブックスライブラリープロジェクトなどをつうじてオフラインデータを、Google+などをつうじてSNSデータを、そしてアンドロイド（Android）[注49]機器をつうじてディバイスのデータまでも収集している。

このように最も多いデータと最も多様なデータを

表7. ビッグデータの社会・経済的意味

区　　分	主　な　内　容
・天然資源 （Natural resources: the new oil, goldrush and of course data mining）	・データに含まれる価値と可能性の注目 ・社会的に懸案とリスクを解決できる潜在力への期待 ・新たな経済的価値の源泉として活用
・自然災難 （Natural disasters: data tornado, data deluge, data tidal wave）	・情報の氾濫により機会を把握と規定遵守が困難 ・増えるデータにより現状態を維持するのに予算が使われ、革新のための新しい投資が困難 ・データ処理の低い応答速度が企業の生産性低下に繋がる恐れ
・産業的道具 （Industrial devices: data exhaust, firehose, Industrial Revolution）	・データの効率的管理と分析をつうじて企業の競争優位の確保 ・データを迅速に処理し、リアルタイムな意思決定を支援 ・データ分析能力が企業の競争力を左右

出所：Tyler Bel. Big Data: An opportunity in search of a metaphor（2011）.

表8. ビッグデータの経済的価値の展望

機　　関	主　な　内　容
・Economist(2010)	・ビッグデータは資本や労働力とほぼ同等な水準の経済的投入資本、ビジネスの新しい原資材の役割 ・ビジネストレンドの把握、疾病予防、犯罪解決などの効果
・MIT Sloan(2010)	・ビッグデータの分析･活用できる組織ほど差別的競争力と高い成果の創出 ・組織分析キャパシティの特徴を提示
・PwC(2010)	・ビッグデータは今まで不可能であったデータの活用を可能にし、潜在的価値と影響力が高い ・ビッグデータの重要性について企業が注目し、新しいビジネスの価値創出のコアとなる
・Gartner(2011)	・データ（情報）は21世紀の石油、データが未来競争力を左右 ・企業は来るデータ経済時代を理解し、情報孤立（information silos）を警戒すべき ・ビッグデータは今後注目すべきエマージング技術
・McKinsey(2011)	・グローバルビジネス環境を変える技術トレンドの3つのコアはクラウド、ビッグデータ、スマート資産である ・ビッグデータは革新、競争力、生産性のコア要素 ・医療、公共行政など5大分野で6千億ドル以上の価値創出

出所：ズン・ジスン(2011)、p.14。

表9.　日本におけるビッグデータの活用を推進するための具体的な方策

具体的な方策	今後の推進に向けたアクション
・官民のデータのオープン化･横断的利活用が可能な環境の整備(日本版オープンデータ戦略)	●行政機関や民間事業者等に埋没･散在するデータのオープン化、各種データを社会全体で横断的に利活用することができる環境を整備 ▷2014年度までに、データの二次利用に関するルールを整備 ▷2015年度までに、オープンデータ環境整備に向けた共通APIの 開発および国際標準化を推進
・電気通信事業者における運用データ等の街づくりや防災等への活用に関するガイドラインの策定	●電気通信事業者において保有されている運用データ等について、個人情報等に配慮しつつ活用するための検討の場の設置および街づくりや防災等への活用に関するガイドラインの策定を支援
・多種多量なデータをリアルタイムに収集･伝送･解析等する技術やデータ秘匿化技術等の研究開発･標準化	●多種多量のデータについて、安全性や信頼性を確保しつつ、効率的な収集、リアルタイム解析等を可能とする通信プロトコル、セキュリティ対策、データ構造等に関する研究開発を推進 ●日本が技術的強みを有している物理ネットワーク層（M2M、メッシュNW、センサー、IoT、車車間）の強化（研究開発、標準化） ▷2017年度までに、安全性・信頼性の高いビッグデータ通信規格を開発・実証するとともに、その成果をITU等の国際標準に反映
・ビッグデータ活用人材(技術やビジネス等の様々な分野における知識や能力等を備えた人材)の育成	●高度なデータ解析技術の開発や画期的なデータ活用事例の実証等をつうじた専門家の育成を目指し、競争的資金の活用を推進 ●JGN-Xを用いたビッグデータ解析基盤の構築および若手研究者やベンチャーへの開放
・安全性・信頼性の高いM2Mに関する通信規格の研究開発・標準化	●機器同士が人を介在せずに相互に情報交換し、自動的に最適制御をするための安全性・信頼性の高い通信規格の開発・実証を行い、国際標準化を推進 ●社会実装を目指したM2Mのテストベット環境の構築と技術実証。 ▷2015年度までに、現状の数千倍程度以上のアクセスがあった場合でも支障なくM2M通信の制御を可能とするための基本技術を確立
・ビッグデータの活用に関するICTの利活用を阻む規制･制度改革の促進	●ビッグデータの活用による新サービス創出等に資するICTの利活用を阻む規制・制度改革に関するIT戦略本部を中心とした取組を引き続き促進するとともに、様々な推進体制との連携等により民間ニーズの掘り起こし等を推進
・異業種･産学官の連携によるビッグデータの活用に関する推進体制の整備	●多様な企業・団体・業種の枠を超え、活用可能なデータや成功事例等の共有、活用を阻み得る規制・制度等の課題の抽出、社会受容性やインセンティブの醸成、関連機関への働きかけ等の課題解決に向けた活動等を産学官の連携で推進する場の構築
・外国政府等とのビッグデータの活用に関する対話の強化	●欧米をはじめとする政策動向等に関する定期的な相互対話のための枠組みを引き続き活用
・ビッグデータの活用に関する計測手法の確立	●ビッグデータのデータ量やその活用によりもたらされる経済価値の見える化等のための計測手法を開発 ▷ 2013年度中に、調査手法および評価手法の確立

出所：情報通信審議会ICT基本戦略ボード. ビッグデータの活用の在り方について. ビッグデータの活用に関するアドホックグループ取りまとめ（2012）。

収集しているGoogleはビッグデータ関連技術やツールを開発している。まず、グラフアルゴリズムの処理をサポートするための技術で、1兆個のデータを数秒内で処理することができるPregel（プリジェル）[注50]をあげることができる。次に、大容量のデータを分散処理することで早く分析できる技術としてDremel（ドレメル）[注51]があげられる。これはGoogleFSとBigTableに保存されている巨大なデータ集合のクェリ（条件）を高速化する技術のことで、Hadoop上のMapReduceのバッチジョブで数時間あるいは数日かかる処理も、Dremelではほとんど一瞬で結果が得られる技術である[注52]。

最後に、検索インデックスを作成するために技術として、既存のマップリデュース方法より約100倍迅速に作業を処理[注53]できる大規模データ用の逐次更新処理システムであるPercolator（パーコレーター）[注54]をあげることができる。すなわち、これは数十ペタバイト規模のデータを数千台のマシン上に蓄積し、1日あたり数十億の更新処理を行うシステムである[注55]。

GoogleがICTの世界にもたらした新しいパラダイムをまとめてみると次のようである[注56]。すなわち、第1に、増え続ける巨大容量の非定型データ、第2に、深い分析、第3に、分散並列処理を用いた高速処理、第4に、コモディティハードウェアの利用、そして第5に、スケーラビリティ、などである。

2）Amazonのビッグデータ戦略

「地球上で最も顧客を大切にする企業である」と表明するAmazon.com[注57、注58]は、米国ワシントン州シアトルを本拠地とするフォーチュン500社の一企業であり、eコマースにおける世界的なリーディングカンパニーでもある。ジェフ・ベゾス（Jeff Bezos）が1995年に設立して以来、Amazon.comは商品の品揃え、インターナショナル・サイト、そして世界中に位置する物流センターおよびカスタマーサービスセンターにおいて著しい拡大を行ってきた。現在では、書籍、エレクトロニクス製品からテニスラケット、宝飾品まで様々な商品を豊富に取り揃え[注59]、2014年10月現在、アメリカ、イギリス、ドイツ、フランス、日本、カナダ、イタリア、中国、インド、スペイン、メキシコ、ブラジルなどの12カ国でウェブサイトを運営し、世界各地50カ所を超える物流センターを設置・運営している。

そして、テクノロジーの進歩はAmazon.comの急速な発展を促し、より多くの商品をより便利に、さらに低価格で顧客に提供することを可能にした。顧客用にカスタマイズされたショッピング体験、「なか見！検索」機能による書籍の検索、「1-Click Shopping」を使った代金の支払い、またリストマニアやほしい物リストなど顧客のショッピングをサポートするコミュニティ機能などを提供している。

Amazonは自社のウェブサイトで商品を購入した顧客の購買内訳をデータベースに記録・貯蔵し、このデータの分析により顧客の消費トレンドと関心事などの消費パターンを把握する。そして、ビッグデータの分析・活用をつうじて顧客別にレコメンデーション（recommendation、推薦商品）を表示する[注60]。この競合他社が真似できない、最も競争力の高いレコメンデーション機能はA9[注61]といわれる検索エンジンによって行われ、過去の購入履歴などから顧客一人ひとりの趣味や読書傾向を探り出し、それに合致すると思われる商品をメール、ウェブページ上で重点的に顧客一人ひとりに推奨する機能のことである。たとえば、Amazon.co.jpの「トップページ」や「おすすめ商品」では、そのユーザーが、過去に購入したり閲覧した商品と似た属性を持つ商品のリストがレコメンデーション機能の一部として自動的に提示される。シリーズ物の漫画などの購入をレコメンドする場合にはちょうど新刊が出た頃に推奨し、似たような傾向の作品をも推薦する。

また、最近にはFacebook情報と連携し、ユーザーの知り合いが購買、また欲する商品を推薦する機能も提供している。

3）Facebookのビッグデータ戦略

周知のごとく、Facebook（フェイスブック）は2014年6月現在、全世界で13億2千万人以上の活動ユーザーをもつ世界最大のソーシャルネットワーキングサービス（SNS）である。すなわち、ユーザーが互いの個人情報と文書、動画などを相互交流するSNSの代表的なもので、それ自体でクラウドであり、ビッグデータプラットフォームあるともいえる。

Facebookは個人のプライベット情報や関心事、活動内訳などのさまざまなデータをインターネットのみならず、オフラインをつうじても絶えず収集・分析し、これを広告に活用することで収益を創出している。そして、内部組織のプロセス分析にもビッグデータ技術を積極的に活用し、Facebookに自社の職員が投稿する文書やタイムラインなどを分析し、互いにコミュニケーションが活発な職員をチームとして構成するなど組織力向上にもビッグデータを

活用している[注62]。

Ⅵ. むすびに ービッグデータ活用の課題ー

以上、述べてきたように、ビッグデータは医療・保険、公共部門、マーケティング、製造業などの分野で活用することにより経済的価値の創出することができる。

このようなビッグデータとは、通常のデータベース管理ツールが貯蔵・管理および分析可能な範囲を超える規模のデータと定義づけることができる。すなわち、ビッグデータを既存のシステム、サービス、企業などで与えられた費用や時間で処理・分析できるデータの範囲を超えるデータのことである。

本稿では、ビッグデータの登場背景と概念的考察を踏まえ、ビッグデータの市場状況と活用のための技術、そしてGoogle、Amazon、Facebookなどのビッグデータの活用事例の考察を行ってきた。

ここでは、ビッグデータの活用における主な課題[注63]を述べることでむすびとしたい。まず第1に、データ品質の確保課題をあげることができる。ビッグデータの品質確保の観点からエンドユーザーのデータの品質を次の4つに類型化[注64]することができる。すなわち、①内在的データ品質（Intrinsic DQ; 正確性、客観性、信頼性、評判）、②接近性（アプローチ）データ品質（Accessibility; アプローチ、アプローチ・セキュリティ）、③文脈（状況）的データ品質（Contextual DQ; 関連性、付加価値、適時性、完全性、データの量）、そして④表現的データ品質（Representational DQ; 解析力、理解しやすさ、簡潔性、表現の一貫性）などである。ビッグデータの利活用におけるデータ品質を確保するためには上記の4つのカテゴリーと細部品質要素をすべて満足する場合、データの品質が保証されるといえる。

したがって、企業や組織においてビッグデータを利活用するためには上記の示した4つのカテゴリーを中心としたガイドラインを構築し、組織内で体系的に管理する必要性がある。

第2に、個人情報保護とプライバシー侵害の問題をあげることができる。ビッグデータの活用において最も大きな危険要素として個人情報保護とプライバシー侵害の問題[注65]をあげることができる。すなわち、過去には考えられないほどの大きさや速度をもつビッグデータの活用において、個人情報の侵害可能性と危険性が一層高くなっている。とくに、最近、頻発に発生している個人情報の流出事故[注66]によりビッグデータの活用において個人情報保護に関する細部的な法制度の改正など具体的な対応の検討などが至急に要求される。

しかしながら、ビッグデータの活用による新たな価値創出は市場の経済的効果や生産性の向上をもたらすことで、産業界などから規制緩和の声も大きい。したがって、個人情報保護と規制緩和の調和を持続的に模索する必要がある。

最後に、専門人材の育成の課題をあげることができる。ビッグデータ時代において、ビッグデータの活用の際に、ビッグデータの技術インフラの開発とデータサイエンティストの人材育成は不可欠な要件である。すなわち、ビッグデータの優秀な人材のもつスキルはビッグデータの活用において成功をも左右するほど重要なファクターである。

しかし、McKinseyによると、アメリカでは、2018年まで14万から19万名のビッグデータ専門家および150万名程度のデータ管理者と分析人材が不足すると予想している。

これを踏まえて、アメリカとヨーロッパ諸国の大学はビッグデータ関連専門人材の養成のための実務型専門課程を運用し、主に産業工学、統計学、コンピュータ工学など関連学部間の協力によるプログラムが進行されている。また、学際間の融合研究をつうじてデータ分析アルゴリズム、機械学習、人間基盤研究、複雑系システムのモデリング、数理的シミュレーション技法、並行プログラミングなど多様な研究が進行されている。

以上のように、ビッグデータについて述べてきたが、最後に一言、ビッグデータはそれ自体では何の意味合いをも持たない。その脈絡と解析をどこから求めるべきかが重要である。これは今後の大きな課題であり、注目すべきところである。ビッグデータの今後の動きから目を離すことはできない。

注

注1 The Economist (2010), The data deluge.

注2 アルビン・トフラー（Alvin Toffler）は『未来の衝撃（Future Shock）』をつうじて情報の洪水（Deluge of Information）という概念を一般化させた。

注3 1億人以上のアクティブユーザーを有するツイッターは1日あたり2億5000万ものつぶやきが発生し、ツイッター全体でみれば1日に8テラバイトものデータが生み出されている。また、Facebookは毎日25億件のコンテンツ、500テラバイト以上のデータを処理している。また「いいね」は毎日27億回、アップロードされる写真も毎日3億枚で、30分ごとに105テラバイトのデータがスキャンされている。Googleは、1日に24ペタバイト以上のデータを処理していると言われている（BELINDAのウェブサイト資料）。

注4 ビッグデータの概念は、2001年には登場していた。一言でいえば、ビッグデータとはみんなのデータのことである（ZDNet JAPAN）。

注5 Big DataはVery large data, Extreme data, Total dataなどと呼ばれている。

注6 McKinsey&Company.http://www.mckinsey.com/（2014-12-14）.

注7 非構造化データとは、既存企業の売上額、個人の年齢や性別などの構造化されたデータではなく、テキスト、音声、文字メッセージなど多様な類型のデータのことをいう。

注8 M2Mとは、Machine to Machine、すなわち、機器間の通信を意味し、人間の介在なしに機器同士がコミュニケーションし動作するシステムのことである。人間同士の通信をH2H（Human to Human）、人と機器との通信をH2M（Human to Machine）、M2H（Machine to Human）などもある。

注9 IoT（Internet of Things）とは、コンピュータなどの情報・通信機器だけでなく、世の中に存在するさまざまなものに通信機能を持たせ、インターネットに接続したり相互に通信することにより、自動認識や自動制御、遠隔計測などを行うことをいう。たとえば、自動車の位置情報をリアルタイムに集約して渋滞情報を配信するシステムや、人間の検針員に代わって電力メーターが電力会社と通信して電力使用量を申告するスマートメーター、大型の機械などにセンサーと通信機能を内蔵して稼働状況や故障箇所、交換が必要な部品などを製造元がリアルタイムに把握できるシステムなどが考案されている（IT用語辞典e-Words）。

注10 これについては、第Ⅳの1を参照されたい。

注11 クラウドコンピューティングとは、共用の構成可能なコンピューティングリソースス（ネットワーク、サーバー、ストレージ、アプリケーション、サービス）の集積に、どこからでも、簡便に、必要に応じて、ネットワーク経由でアクセスすることを可能とするモデルであり、最小限の利用手続き、またはサービスプロバイダとのやりとりで速やかに割当てられ提供されるものである（Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.米国立標準技術研究所「NISTによるクラウドコンピューティングの定義」p.1）。

注12 Amazon Elastic Compute Cloud（Amazon EC2）とは、規模の変更が可能なコンピュータ処理能力をクラウド内で提供するウェブサービスである。

注13 Amazon S3（Simple Storage Service）は、Amazon Web Serviceが2006年に開始したクラウドストレージサービスで、Webサイトのホスティング、イメージやビデオのホスティング、バックアップストレージをはじめ、広範囲な用途に利用されている。

注14 海部（2013）、p.19-20。

注15 ビッグデータという概念は、1997年ACMのProceeding of the 8th conference on VisualizationでMichael CoxとDavid Ellsworthの論文「Application-controlled demand paging for out-of-core visualization」で初めて発表された。

注16 Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis (IDC, 2011, p.6).

注17 ビッグデータとは何か. http://semi.miyazaki-mu.ac.jp/~skaneko/tkaneko/bigdata（2014-12-14）.

注18 総務省a（2012）、p.153。

注19 野村総合研究所. ビッグデータ時代の到来. ITソリューションフロンティア29-3, 6（2012）。

注20 オラクル：エンタープライズ向けのビッグデータ. Oracleホワイト・ペーパー、p.3。

注21 成者政（2007）、p.122-142。

注22 CDRと呼ばれる詳細な通話記録のことである。

注23 構造化データと非構造化データ. http://www.amy.hi-ho.ne.jp/kido/kouzouka.htm（2014-12-14）.

注24 横山ほか（2012）、p.13。

注25 網野（2013）、p.24。

注26 ベ・ドンミンほか（2013）、p.41。

注27 karikaho.ところで、ビッグデータって何？～Veracity（正確）. http://ameblo.jp/karikaho/entry-11391625443.html/（2014-12-14）.

注28 Lee, D.J（2013）, p.11.

注29 XaaS（X as a Service）とは、情報システムの構築・運用に必要な何らかの資源（ハードウェア、回線、ソフトウェア実行環境、アプリケーションソフト、開発環境など）をインターネットをつうじてサービスとして遠隔から利用できるようにしたもの、また、そのようなサービスや事業モデルのことである。従来は購入したり固定的・長期的な利用契約を結んで利用したさまざまな資源を、サービスとしてネットワーク越しに必要なときに必要なだけ利用し、実績に応じて代金を支払う形態を意味する。「サービスとしてのソフトウェア」（SaaS：Software as a Service）の概念を広げ、さまざまな要素に適用できるよう

にした用語である。XaaSに含まれる概念には、SaaSのほかに、ソフトウェア実行環境を提供するPaaS（Platform as a Service）や、仮想化されたサーバーや回線などのハードウェア環境を提供するIaaS（Infrastructure as a Service）またはHaaS（Hardware as a Service）などがある（IT用語辞典e-Wordのウェブサイト資料）。

注30 ビッグデータの活用に関する技術については、城田（2012）、p.47-82を参照されたい。

注31 ハドゥープについての詳細は、太田一樹ほか. Hadoop徹底入門. 翔泳社（2013）；佐々木達也. Hadoopファーストガイド. 秀和システム（2012）；田澤孝之. 基礎から解説！企業を変えるHadoop. ITproのウェブサイト資料などを参照すること。

注32 What Is Apache Hadoop?. http://hadoop.apache.org/（2014-12-14）.

注33 オープンソースソフトウェア（OSS）とは、ソフトウェアの設計図にあたるソースコードを無償で公開し、誰でもそのソフトウェアの改良、再配布が行えるようにすること。そしてそのようなソフトウェアのことを指す。

注34 石井一夫. 医療、農学、環境分野におけるビッグデータ解析. 生物工学会誌92-2. 日本生物工学会（2014）、p.92。

注35 Davenport（2014）, p.115.

注36 Googleが独自に開発した分散ファイルシステムで、Googleのクローラが集めてきた大量のコンテンツ、Gmailなどの利用者が保存する大量のファイル、Google Mapsなどが表示する大量の画像など、想像を絶するほど大量のデータが保存されている（Googleが「Google File System」次期バージョンを開発中）。http://www.publickey1.jp/blog/09/google_file_system.html/（2014-12-14）.

注37 HDFS（Hadoop Distributed File System）とは、分散バッチ処理ソフト「Apache Hadoop」向けのファイルシステムである。これはファイルを分割して複数のディスクで管理することで、大量データ処理のスループットを引き上げる役割をする（森山 徹. Hadoopを支える「HDFS」. http://itpro.nikkeibp.co.jp/article/Active/20120912/422326/）。

注38 稲田（2012）、p.200。

注39 NoSQLについては、本橋信也ほか. NOSQLの基本知識（ビッグデータを活かすデータベース技術）. リックテレコム（2012）；佐々木達也. NoSQLデータベースファーストガイド. 秀和システム（2011）などを参照すること。

注40 森川博之. ビッグデータの活用に関するアドホックグループの検討状況（2012）、p.13。

注41 データウェアハウスについては、鈴木憲司. データウェアハウスがわかる本. オーム社（2000）；データウェアハウス研究会. よくわかるデータウェアハウス. 日本実業出版社（2000）；安達敏光、TERADATAのウェブサイト資料などを参照されたい。

注42 Inmon（2005）, p.29-33.

注43 安達敏光. データウェアハウスの中身と効用. http://jpn.teradata.jp/library/nyumon/ins_1905.html /（2014/12/14）.

注44 McKinsey Global Institute（2011）, p.12.

注45 栗田（2013）、p.3-4。

注46 財団法人日本情報処理開発協会. パーソナル情報の利用のための調査研究報告書（2011）、p.124。

注47 高度情報通信ネットワーク社会推進戦略本部（2014）、p.1。

注48 Googleの共同創設者で CEO のラリー ペイジは「完璧な検索エンジンとは、ユーザーの意図を正確に把握し、ユーザーのニーズにぴったり一致するものを返すエンジンである」と発言している。

注49 これはスマートフォンやタブレットなどの携帯情報端末のために開発されたプラットフォームであるが、「多種多様なディバイスにネットワーク接続機能を具備させるための閾値を一気に下げたソフトウェア」として捉えるべきである（鈴木（2011）、p.30）。

注50 これについては、Grzegorz Malewicz et. al., A System for Large-Scale Graph Processing, p.135-145を参照されたい。

注51 これについては、Sergey Melnik et. al., Dremel: Interactive Analysis of Web-Scale Datasets, Proceedings of the VLDB Endowment, Vol.3, No.1, 2010を参照されたい。

注52 TechCrunch Japanのウェブサイト資料. ApacheがGoogleのリアルタイムビッグデータツールDremelのオープンソースクローンDrillを（2014-12-15）による。

注53 鈴木（2011）、p.168。

注54 詳細については、Daniel Peng and Frank Dabek, Large-scale Incremental Processing Using Distributed Transactions and Notifications, p.1-14.を参照されたい。

注55 丸山不二夫. 大規模分散システムの現在. http://www.slideshare.net/maruyama097/large-system（2014-12-14）.

注56 （独）情報処理推進機構（2012）、p.8。

注57 Amazonの概要とビジネスモデルなどについては、成者政（2006）、p.43-45を参照されたい。

注58 About Amazon. http://www.amazon.co.jp/version2/b/ref=footer_about?ie=UTF8&=52267051（2014-12-14）.

注59 2014年10月現在のカテゴリーとしては、Kindle本＆電子書籍リーダー、Fireタブレット、Amazonインスタントビデオ、デジタルミュージック、Amazon Cloud Drive、Android アプリストア、ゲーム&PCソフトダウンロード、本・コミック・雑誌、DVD・ミュージック・ゲーム、家電・カメラ・AV機器、パソコン・オフィス用品、ホーム＆キッチン・ペット、食品＆飲料、ヘルス&ビューティー、ベビー・おもちゃ・ホビー、ファッション・バッグ・腕時計、スポーツ＆アウトドア、そしてDIY・カー＆バイク用品などである。

注60 この機能により全体Amazonの売り上げの30%を占めている。

注61 A9は2003年に設立した検索関連技術を保有する企業で、Amazonの商品検索、書籍本文検索、そしてイメージ検索サービスなどを提供する子会社である。

注62 リュ・ハンソク. ビッグデータビジネスの争点と展望. datamining.dongguk.ac.k（2014-12-14）.

注63 野村総合研究所が2012年7月に実施した「ビッ

グデータの利活用に関するアンケート調査」の結果によると、「今後、貴社でビッグデータを進めていく場合、どのようなことが問題・課題となりそうですか」という設問に対し、「ビジネスとして具体的に何に活用するかが明確ではない（61%）」、「投資対効果の説明が難しい（45%）」、「担当者のスキルが不足している（45%）」、「ビジネスとデータ分析の両視点で戦略を考えられる人材がいない（36%）」、「担当者の人数が不足している（32%）」、そして「ビッグデータ活用の受け皿になる組織がない（29%）」などの回答（複数回答）を得ている。

注64　Strong et. al.（1997）, p.103-110.

注65　個人情報保護の一般的な概要については、成・葛西（2000）を参照されたい。

注66　ベネッセコーポレーションは2014年7月9日、同社の顧客情報約760万件が外部に漏えいしたことを確認したと発表した。顧客情報が漏えいしたデータベースに保管されている情報の件数から推定すると、最大で約2,070万件の顧客情報漏えいの可能性があるという（ScanNetSecurutyのウェブサイト資料）。

文献

1) 網野知博. 会社を強くするビッグデータ活用入門. 日本能率協会マネジメントセンター(2013)
2) 城田真琴. ビッグデータの衝撃. 東洋経済新報社(2012)
3) 稲田修一. ビッグデータがビジネスを変える. アスキー・メディアワークス(2012)
4) 横山隆治ほか. ビッグデータ時代の新マーケティング思考. ソフトバンククリエイティブ株式会社(2012)
5) 鈴木良介. ビッグデータビジネスの時代. 翔泳社(2011)
6) 海部美知. ビッグデータの覇者たち. 講談社(2013)
7) (独)情報処理推進機構. つながるITがもたらす豊かなくらしと経済(2012)
8) 長橋賢吾. ビッグデータ戦略：大規模データ分析の技術とビジネスへの活用. 秀和システム(2012)
9) 栗田克之. ビッグデータ利活用におけるプライバシー情報の経済的価値と問題解決に向けて. 情報通信学会第30回学会大会個人研究発表資料(2013)
10) 高度情報通信ネットワーク社会推進戦略本部. 世界最先端IT国家創造宣言工程表(2014)
11) 総務省a. 平成24年版情報通信白書(2012)
12) 総務省. 平成25年版情報通信白書(2013)
13) 総務省. 平成26年版情報通信白書(2014)
14) 成者政. CRMによる戦略的企業経営管理. 船越克己ほか編. 企業の経営を支える情報･意思伝達システム. 創成社(2007)
15) 成者政. インターネットショッピングモール企業におけるCRM戦略の構築. 朝日大学経営学会. 経営論集21, 35-54(2006)
16) 成者政・葛西和廣. 高度知識情報化社会における個人情報保護に関する考察－個人情報の収集・侵害類型と侵害防止技術を中心に－. 松本大学. 松本大学研究紀要 8, 49-68(2000)
17) ITR White Paper. データ活用サイクルの重要性～ビッグデータ時代に求められるデータ活用基盤とは～. 株式会社アイ・ティ・アール(2014)
18) ベ・ドンミンほか. ビッグデータの動向および政策の示唆点. 情報通信政策研究院. 情報通信放送政策25-10(2013)
19) ズン・ジソン. 新しい価値創出エンジン、ビッグデータの新しい可能性と対応戦略. NIA. IT & Future Strategy18(2011)
20) OECD. EXPLORING DATA-DRIVEN INNOVATION AS A NEW SOURCE OF GROWTH: Mapping the Policy Issues Raised by Big Data(2013)
21) Thomas H. Davenport. big data@work: Dispelling the Myths, Uncovering the Opportunities. Harvard Business School Publishing Corporation(2014)
22) Tyler Bell. Big Data: An opportunity in search of a metaphor(2011)
23) Bill Franks. Taming The Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics. Wiley (2012)
24) IDC. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east(2012)
25) IDC. Extracting Value from Chaos(2011)
26) McKinsey Global Institute. Big data: The next frontier for innovation, competition, and productivity（2011）
27) McKinsey Global Institute. Game changers: Five opportunities for US growth and renewal (2013)
28) Lee D.J. A Study on the SMEs' Marketing Implementation Using Big Data. KOSBI (2013)
29) William H. Inmon. Building the Data Warehouse(4th.ed.）. Wiley Publishing, Inc. (2005)
30) Paul Zikopoulos et. al. Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw Hill(2012)
31) Diane M. Strong et. al. Data Quality in Context-A study reveals businesses are defining data quality with the consumer in mind-. Communication of the ACM 40-5 (1997)