

### 3. 公的統計におけるビッグ・データの活用の動向

本章では、ビッグ・データの動向、概念、意義、海外の政策動向および公的統計におけるビッグ・データ活用の取り組みについて、紹介する。

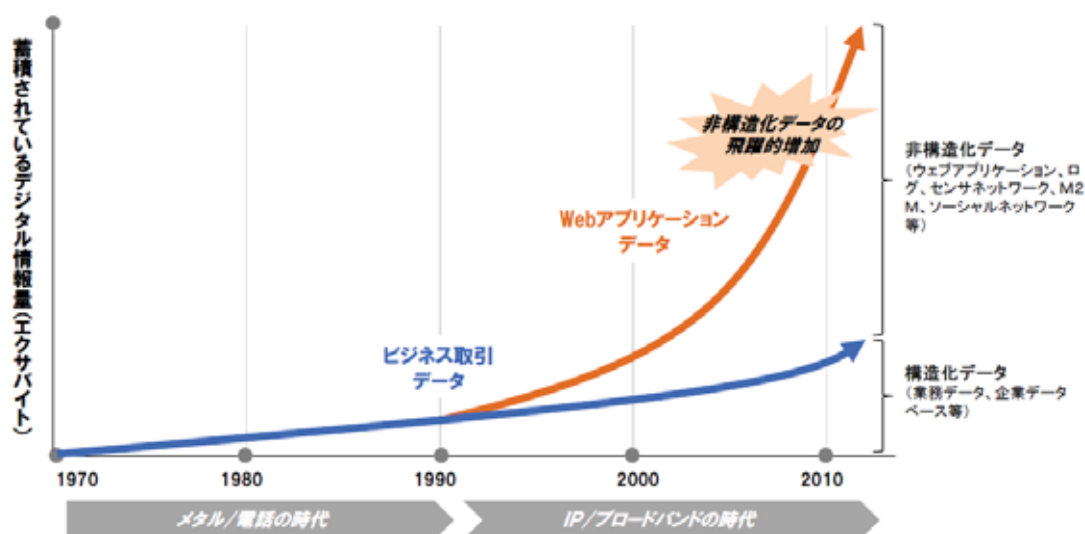
#### 3.1 ビッグ・データの動向

##### 3.1.1 構造化データと非構造化データの伸び

インターネットの普及、コンピュータの処理速度の向上や無線通信技術の進展、さらに、スマートフォン等の多様なデバイスの普及によって、ビッグ・データと呼ばれる膨大なデジタルデータが蓄積している。

ビッグ・データは、これまで蓄積されてきた企業内で管理する顧客データベースや業務データなどの「構造化されたデータ」と、構造化されていない多種・多量なデータ、たとえば音声、映像やソーシャルメディア、さらにセンサーから取得された「非構造化データ」に分けられ、とりわけ後者において、生成・蓄積・流通されるデータ量が飛躍的に拡大している（図表 3-1）。

図表 3-1 構造化データと非構造化データの伸び（イメージ）



(出典) 総務省「ビッグデータ時代における情報量の計測に係る調査研究報告書」(2014年3月)

##### 3.1.2 ビッグ・データの種類

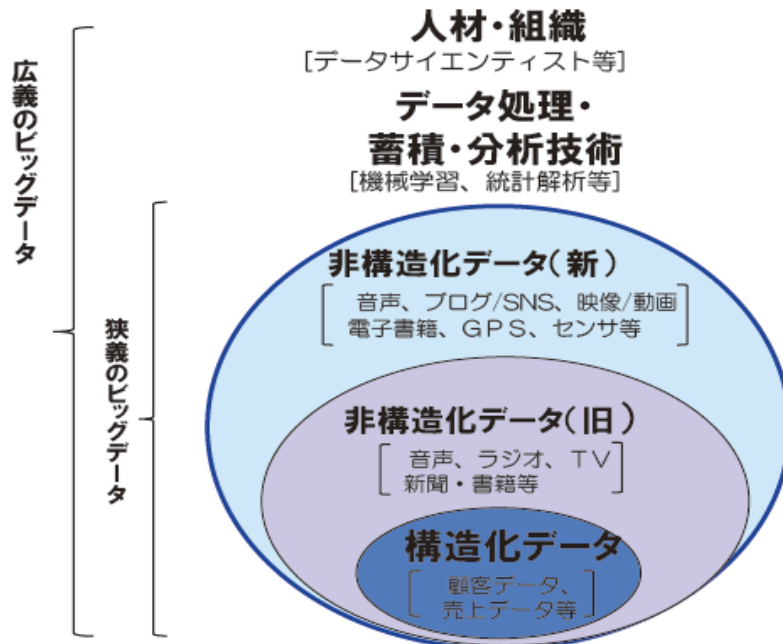
###### (1) ビッグ・データの種類

ビッグ・データの種類は、狭義のビッグ・データ、広義のビッグ・データに分けられる。

狭義では顧客データ、売上データ等の「構造化データ」、従来からあった「非構造化データ(旧)」(音声、ラジオ、TV、新聞、書籍等)そして新たなデバイスの登場によ

って爆発的に増えている「非構造化データ(新)」(音声、ブログ/SNS、映像/動画、電子書籍、GPS、センサ等)までが含まれるが、広い概念としては、ビッグ・データを扱う「データ処理・蓄積・分析技術(機械学習、統計学習等)」や、人材・組織(データサイエンティスト等)までを含んでいる(図表 3-2)。

図表 3-2 ビッグ・データ概念



(出典) 総務省「情報流通・蓄積量の計測手法の検討に係る調査研究」(平成 25 年)

(2) ビッグ・データ特性

ビッグ・データ特性は、以下に示すように Volume, Variety, Velocity の 3V で表される<sup>1)</sup>。

- Volume(ボリューム:量)

第 1 の特性は、データ量が圧倒的に多いことがあげられる。企業内部で増加するデータは、従来のトランザクションによるものや既存データと同様に、新たなタイプのデータがあり、ストレージの課題、大規模分析の課題があがっている。

米調査会社 IDC によると、2012 年に 2.8 ゼタ (1 兆の 10 億倍) バイトだった全世界のデータ量は、20 年には 14 倍の 40 ゼタバイトに膨らむ見通しである。

<sup>1)</sup>最近では Value (価値) を加えて 4V、Veracity(情報の正確さ,信憑性) を加えて 5V ともいわれている。

- **Variety(バラエティ：多様性)**

第2の特性は、さまざまなデータである。従来から企業内に存在する販売データや在庫データなどに加えて、主として、ソーシャルメディアやモバイルから生成されたデータなどが多様に広がっている。

データの種類は、表形式、階層型、ドキュメント、メール、ツイッターやフェイスブックなどソーシャルメディア内のテキストデータ、携帯電話・スマートフォンのGPSから発生する位置情報、時々刻々と生成されるセンサーデータ、ビデオ、静止画、オーディオ、株価、決済データなどさまざまである。

- **Velocity (ベロシティ：速度)**

第3の特性は、ビッグ・データを入出力できる速度である。速度とは、データの生成の速度と同時に、それをどれだけ素早く処理しなければならないかという要求の速度、更新頻度の高さを意味している<sup>2</sup>。

### 3.1.3 ビッグ・データ活用の意義

ビッグ・データ活用の意義については、ICTの進展に伴い多種多量なデータの生成・収集・蓄積等をリアルタイムで行うことが可能となり、そのようなデータを分析することで未来の予測や異変の察知等を行い、利用者個々のニーズに即したサービスの提供、業務運営の効率化や新産業の創出等が可能となっている点にある。

マクロ視点では、ビッグ・データを起点としたデータ関連投資・資本ストックの増加やそれに付随した生産性の向上を通じた経済成長が期待される（図表 3-3）。

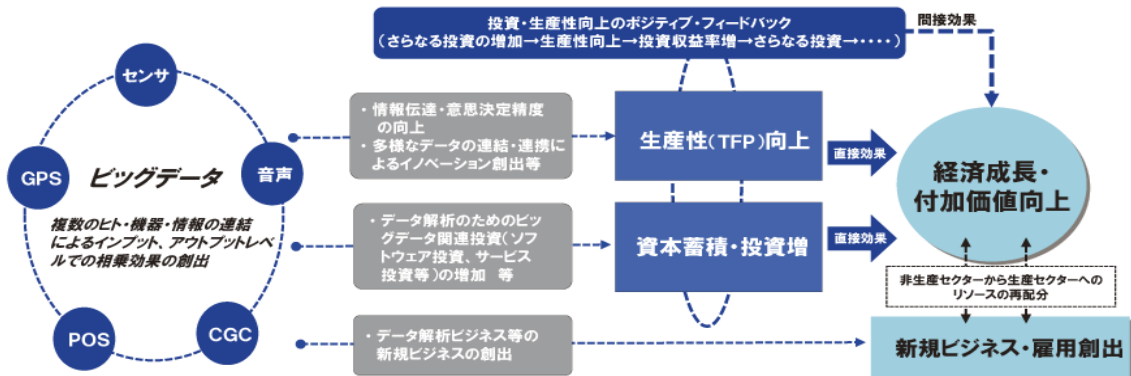
日本においては、各種センサー情報の発達、高性能コンピュータと自然言語処理で世界的に高い水準を保持しており、この強みが幅広い分野・領域に展開することで、我が国の科学技術の共通基盤の強化や産業競争力の強化が期待されている。

あわせて、高速かつ簡単に分析できるテクノロジーも登場し、従来できなかった異変の察知や未来予測、さらに今を描き出すナウキャストイングが高精度に行われ、企業等におけるコスト削減や、新たなビジネスの創造に大きな期待が寄せられている。

---

<sup>2</sup> “Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data : Gartner Special Report Examines How to Leverage Pattern-Based Strategy to Gain Value in Big Data” <http://www.gartner.com/newsroom/id/1731916> (2011.3.27)

図表 3-3 ビッグ・データ活用の意義（マクロ視点<sup>3</sup>）

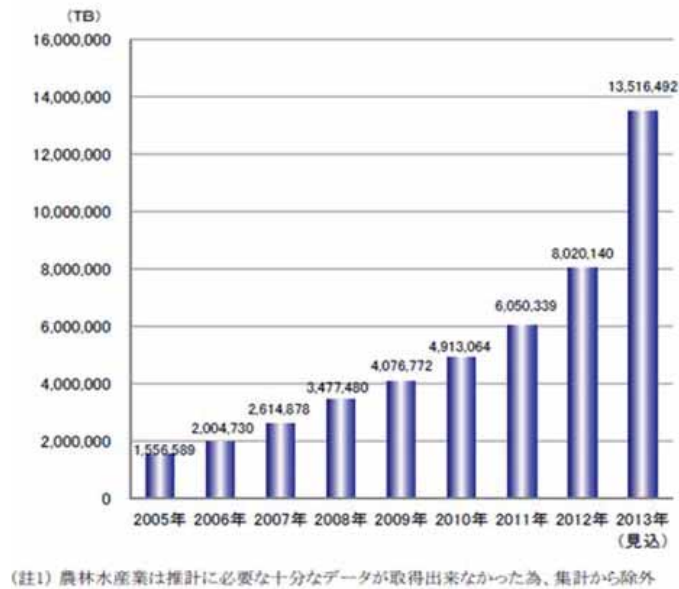


(出典) 総務省「平成26年版情報通信白書」

### 3.1.4 ビッグ・データ流通量の推移

日本におけるビッグ・データ流通量は2005年の1.6エクサバイトから2013年（見込）の13.5エクサバイトへ拡大している（図表 3-4）。

図表 3-4 データ流通量の推移（9産業計）



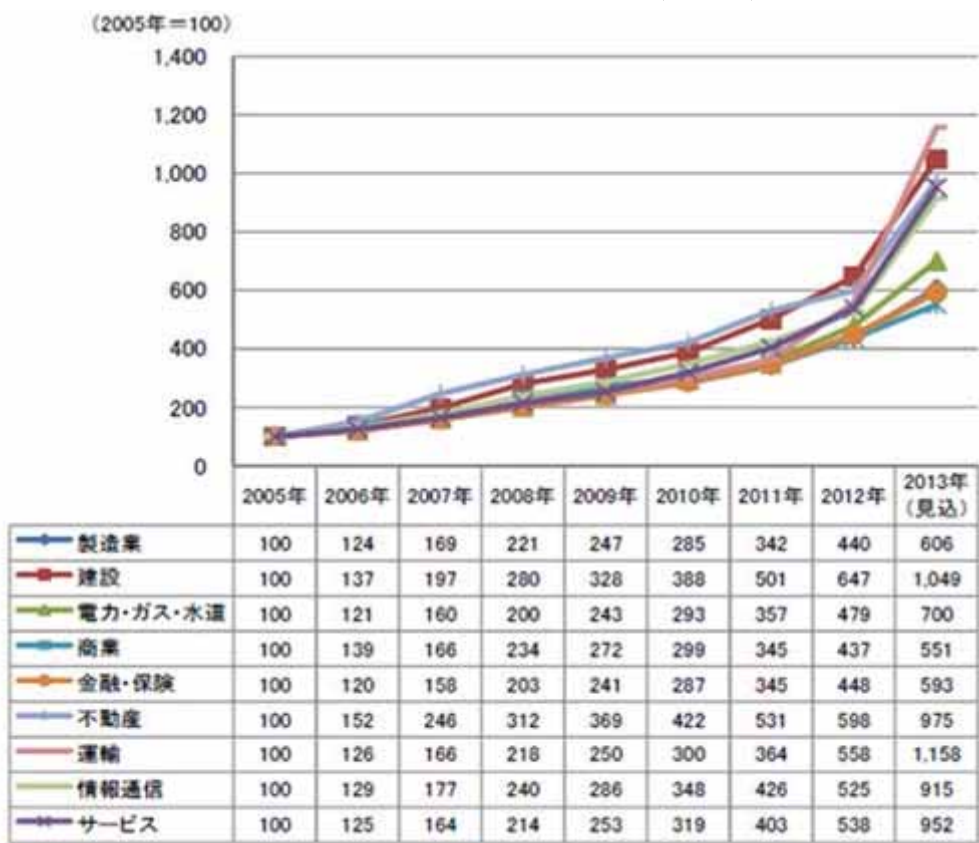
(出典) 総務省「ビッグ・データ時代における情報量の計測に係る調査研究報告書」2014年3月

ビッグ・データ流通量のセクター別推移（2005年=100）をみると、運輸業と建設業の伸

<sup>3</sup>実質 GDP 成長率を労働投入と、資本ストックの投入、技術進歩（全要素生産性上昇）の3つの要因に分解して、それぞれの要因の成長率への貢献を明らかにする成長会計式に当てはめる視点である。

びが顕著である（図表 3-5）。

図表 3-5 データ流通量の推移(業種別)



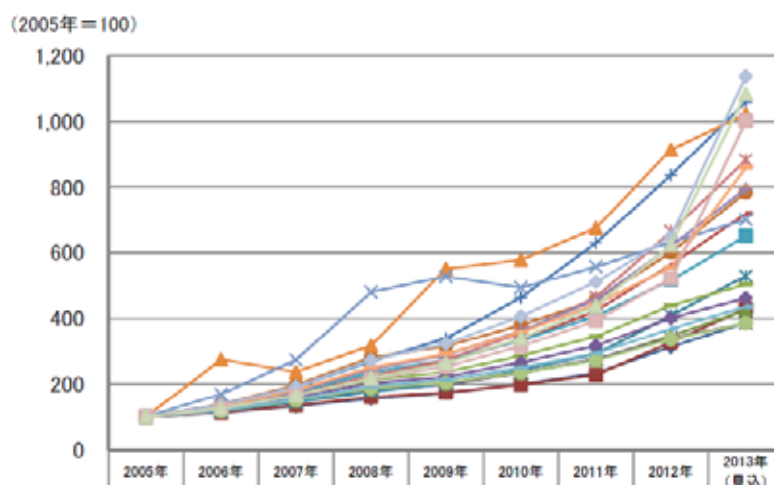
(出典) 総務省「ビッグ・データ時代における情報量の計測に係る調査研究報告書」2014年3月

ビッグ・データ流通量をメディア別にみると、動画・映像視聴ログ、センサーログ、画像診断、GPS データ、交通量・渋滞情報の伸びが大きい（図表 3-6）。

運輸業の情報流通量が大きい背景としては GPS、RFID データ、センサーログ、交通量・渋滞情報量の増大が想定される。

図表 3-6 ビッグ・データ流通量の推移（メディア別内訳）

	2005年	2006年	2007年	2008年	2009年	2010年	2011年	2012年	2013年 (見込)
<b>既存メディア</b>									
顧客DB	100	112	132	155	173	200	233	313	387
経理データ	100	113	135	160	174	199	228	327	430
POSデータ	100	119	149	185	201	234	273	347	425
電子レセプト	100	119	148	181	199	235	276	340	385
業務日誌	100	117	144	177	202	242	291	408	529
電子カルテ	100	137	198	280	315	380	456	601	785
画像診断	100	135	192	268	339	463	630	836	1,059
CTI音声ログデータ	100	125	166	228	287	334	423	564	719
固定IP電話	100	125	164	213	236	286	345	439	504
携帯電話	100	123	160	203	222	265	317	401	482
Eコマースにおける販売ログ	100	128	174	235	273	333	407	517	650
GPSデータ	100	274	237	317	552	578	675	914	1,022
RFIDデータ	100	169	273	480	528	493	558	632	701
気象データ	100	126	168	223	274	356	464	664	882
電子メール	100	119	150	188	204	235	272	337	386
Blog、SNS等記事	100	130	179	244	288	362	456	619	796
アクセスログ	100	120	153	183	214	250	283	369	438
<b>新規メディア</b>									
交通量・渋滞情報	100	131	182	251	291	359	443	556	859
動画・映像視聴ログ	100	135	192	272	323	406	511	649	1,138
防犯・遠隔監視カメラ	100	125	165	216	253	315	393	521	1,002
センサーログ	100	125	166	218	263	340	438	624	1,086



(出典) 総務省「ビッグ・データ時代における情報量の計測に係る調査研究報告書」2014年3月