

ビッグデータの活用の在り方 について

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
取りまとめ

平成24年5月28日
主査 森川 博之

目次

1. ビッグデータの活用に関する背景と現状
… 3
2. ビッグデータの活用に関する国内外の取組と課題
… 22
3. ビッグデータの活用に向けた方向性と具体的方策
… 30

〔別添〕 ビッグデータの活用に関する関係者ヒアリング等の概要

〔参考〕 ビッグデータの活用に関するアドホックグループについて

1. ビッグデータの活用に関する背景と現状

- 国内外の主なシンクタンクによる調査において、いわゆる「ビッグデータ」については、2012年以降のICT(情報通信技術)分野における重要な潮流や戦略的な技術として位置づけられ、関係事業者等において取組が活発化。

Gartner

- 2012年以降にIT部門及びユーザに影響を与える重要な展望「Gartner Predicts 2012」において、次のとおり発表
 - ☞ 2012年に向けて企業が利用できる情報量が増えるものの(「ビッグデータ」、これらを理解することが課題
 - ☞ IT部門によるシステムの管理責任がクラウド等により外部にシフトし、データの一貫性と有効性の確保が困難
 - ☞ 2012年の重要な展望11項目のうち1項目として、「2015年までを通じ、Fortune500企業の85%以上が、ビッグデータを競合優位性確保のために効果的に活用することに失敗」を位置づけ
- 2012年の重要な戦略的技術(今後3年間でITやビジネスに革新を起こすもの、多大な投資の必要が生じるもの等として企業に大きな影響を与える可能性を持つ技術)として、履歴データやリアルタイムデータの分析等へ進化する「次世代アナリティクス」や、膨大な量のデータに関する「ビッグ・データ」を位置づけ



- 2012年における国内IT市場でキーとなる技術や市場トレンド等「Japan IT Market 2012 Top 10 Predictions: 社会基盤を変革する第3のITプラットフォームの台頭」において、主要10項目のうち1項目として、「2012年はビッグデータ活用型アナリティクスビジネスのリーダーの座をかけた競争のスタートダッシュの年になる」を位置づけ



- 今後5年のICT市場のトレンド「ITナビゲーター 2012年版」において、ICT分野における大きな潮流の1つとして、「ビッグデータビジネス(昨今の革新的な情報・通信技術を活用して、きわめて大量のデータを高速で収集・解析することにより、社会・経済の問題解決を図ったり、業務の付加価値を一層高めるための事業)」を位置づけ

- ビッグデータの活用により、将来的に講ずべき施策がわかり、事業を効率的に実施することが可能になるため、事業者においては、例えば、次のような効用が得られることを期待。

製品開発

- ☞ どのような製品を開発することが消費者に対して訴求するのかが分かる。

販売促進

- ☞ 誰に、何を、いつ売れば良いのかが分かる。

保守・メンテナンス・サポート

- ☞ いつ、どのようなメンテナンスを行えばよいか分かる。

コンプライアンス

- ☞ 不正の予兆や、特に注視すべき事象が何であるかが分る。

業務基盤・社会インフラの運用

- ☞ 全般的な性能向上・コスト削減が実現される。

- コンテンツ・アプリケーション、プラットフォーム、ネットワーク及びデバイスの各レイヤーにおけるICTの進展により、多種多量のデータの生成・収集・蓄積等が可能・容易化。

コンテンツ アプリケーション

ソーシャルメディアの普及

- 我が国における直近1年間のソーシャルメディア（mixi、Twitter、Facebook）利用者数は約3200万人となり、ソーシャルメディア消費が1兆5200億円と推計

動画配信サービスの普及

- 平成22年における我が国の動画配信サービスの利用率は約40%が月に数回以上利用し、また、YouTubeは平成23年2月現在で約2900万人が加入

プラットフォーム

クラウドサービスの普及

- 我が国のクラウドネットワーク技術の利用は、2009年度の14.8%から2010年度には22.5%に増加し、クラウドサービスの市場規模は、2015年には約2.3兆円へ成長

Android OSの普及

- 携帯電話端末用OSのみならず、カーナビやデジタルフォトフレーム等の各種機器のOSとして利用され、多種多様なデータの収集等が可能

ネットワーク

ワイヤレス通信の普及

- 電力監視等の遠隔監視分野、電子マネー等の決済データ通信分野等の国内のM2M市場（モバイル回線契約数）は、2015年には、1600万件へ拡大する見込み

ブロードバンド化の進展

- ブロードバンドの基盤整備率は、2011年度末現在、超高速ブロードバンドで約93%に達し、加入契約数は約3500万加入となっており、FTTHが過半数

デバイス

各種センサーの普及

- 加速度センサの世界市場は、コストダウン等により、2013年度に、出荷数量で10億7350万個まで拡大し、市場規模で1675億円となる見込み

ICT端末の多様化

- 世帯保有状況は、2010年末で、携帯電話等が9割、PCが8割を超え、インターネット接続テレビが約3割、スマートフォン及びタブレット型端末が約1割まで普及

センサーの進展

● データの収集等を可能とするセンサーの小型化・低価格化が進展。

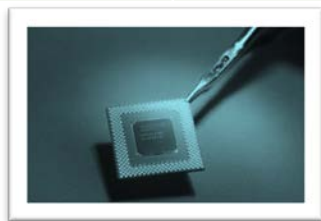
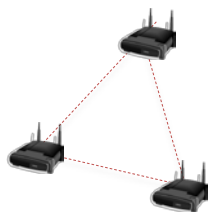
☞ 3軸加速度センサーについて、チップの大きさは2000年の10mm²から2010年の2～3mm²以下へ小型化、平均販売価格は2000年の約240円以上から2010年の約56円程度へ低価格化が進展。

● センサーにより収集等したデータを送信する通信モジュールの低価格化が進展、契約者数も増加。

☞ カーナビや気象観測システム等に搭載される携帯電話の通信モジュールについて、2008年の約2～2.5万円から2010年の約0.6～1万円へ低価格化し、契約数は2008年の約32万件から2010年の約142万件へ増加。

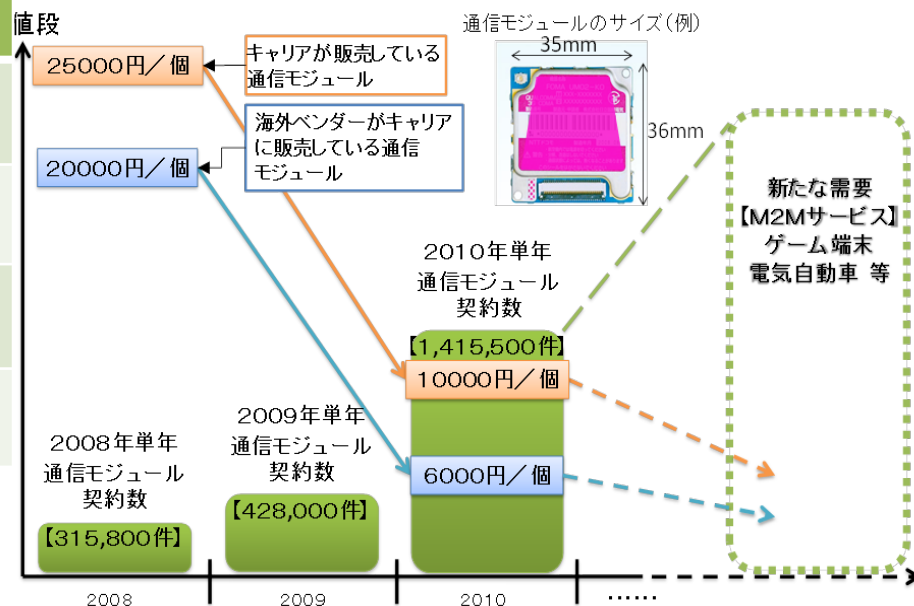
センサの小型化・低価格化

	2000年	2010年	将来 (2020年頃)
チップの大きさ (ダイ表面積)	10mm ²	約2～3mm ²	1～2mm ²
消費電力	0.1mW	0.05mW	0.05mW未満
平均販売価格※	\$3以上 (約240円以上)	\$0.70 (約56円)	\$0.50未満 (約40円未満)
単位生産量	35	771	2500より大



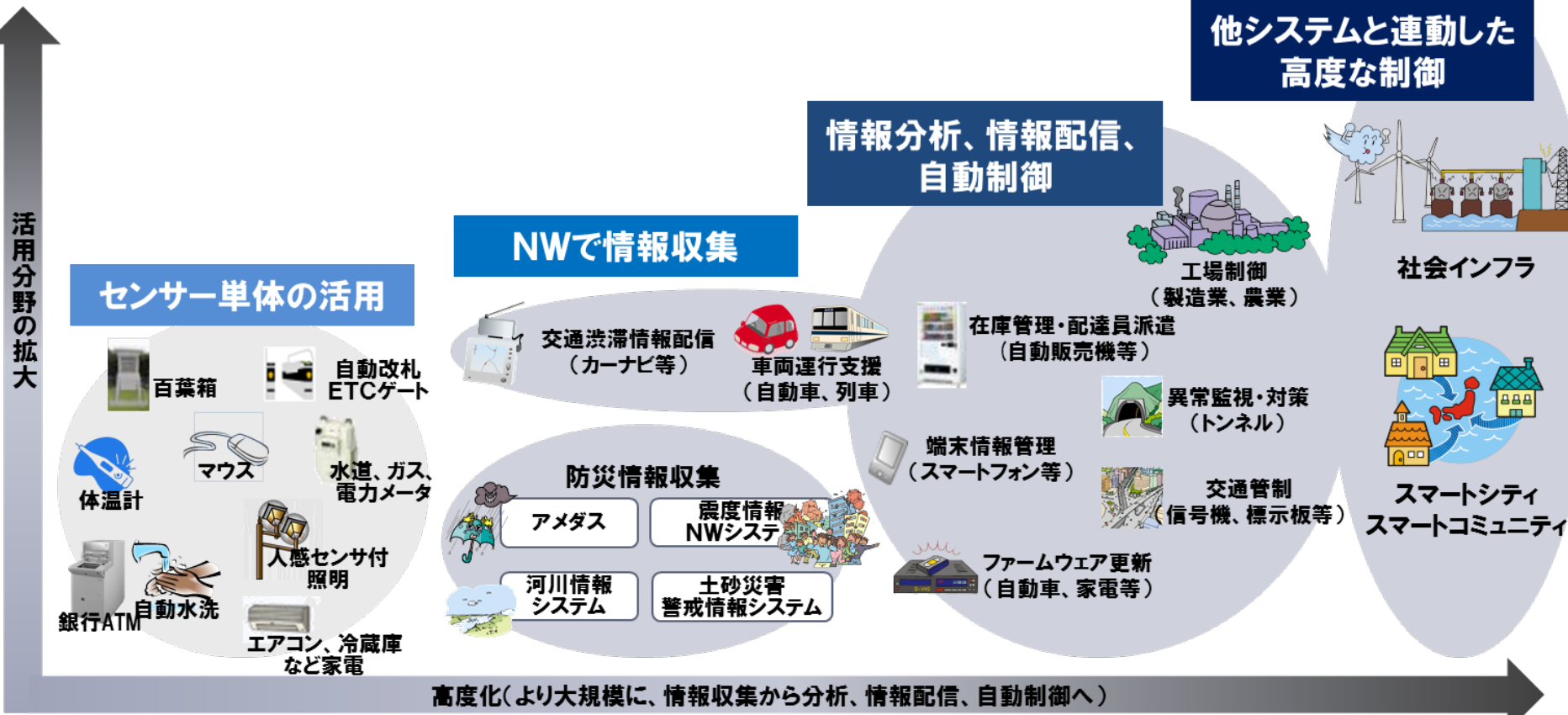
※ 1ドル=80円で換算

携帯電話の通信モジュールの価格推移



センサーネットワークの進化

● センサー単体での活用をはじめとして、現時点では、ネットワークによる情報収集・活用が中心。今後は、情報分析、情報配信、自動制御や他システムと連動した高度な制御へと進展。



- <例> 震度情報の場合
- 震度計の設置** -設置箇所の震度を把握-
 - 震度情報NWシステム** -近隣や全国の震度情報を把握-
 - 緊急地震速報** -地震波の特性を利用し、震源・震度の予報を発表-
 - 緊急地震速報等に基づいた自動制御** -鉄道会社の運行管理システムと連携した列車運行停止などに活用-

【出典】総務省「ICTを活用した街づくりとグローバル展開に関する懇談会」NTTデータ提出資料

M2M通信サービス(例)

● 自動販売機、エレベーター、プラント設備、橋梁等の様々な領域において、M2M通信 (Machine to Machine通信:人が介在せず、ネットワークに繋がれた機器同士が相互に情報交換等を行う機器間通信)サービスが提供。

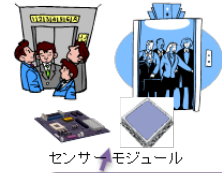
○自動販売機の遠隔モニタリングの例



自動販売機の各種データ収集
(機器の状態、在庫状況、売上状況)

- 販売不能、温度異常、システム異常などの情報に基づく、迅速な故障回復
- 在庫状況に基づく、商品配送のコスト削減
- 売上状況に基づく、マーケティング、販売計画への反映及び廃棄処分品の最小限化

○エレベーターの遠隔モニタリングの例

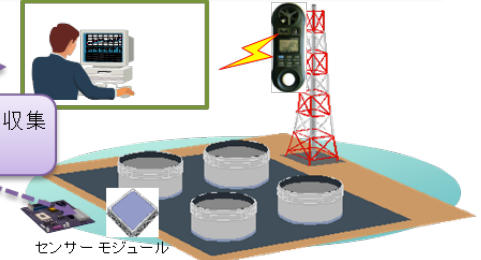


エレベーターの各種データ収集
(機器の状態)



- エレベーターのシステム異常などの情報に基づく、迅速な機器復旧及び人命救助

○プラント設備異常モニタリングの例



工場の操業に関する各種データ収集
(設備の状態、周囲環境)

- 火災検知、有毒ガス漏れ検知、立入禁止区域への侵入者等を検知し、災害や事故を未然に防止

○作業機械の遠隔モニタリングの例

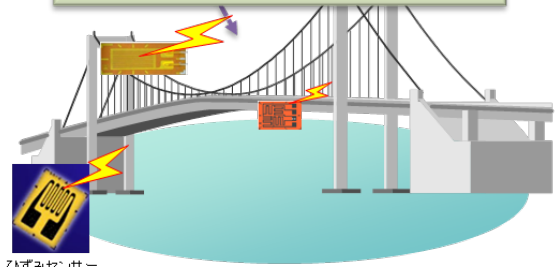


作業機械の各種データ収集
(機器の状態、位置情報、稼働状況)

- 消耗部品の状態や負荷情報に基づく、故障の予防保全
- 運転内容や負荷情報、燃料消費量、CO₂排出量に基づく、省エネ運転支援
- 位置情報、稼働状況に基づく、盗難防止のための遠隔ロック



○構造物劣化モニタリングの例



建物等構造物の異常データを収集
(構造物の状態)

- 構造物のひび割れ、異常な歪み等の危険を検知し、事前のメンテナンスと事故を防止



- 従来の通信サービスとM2M通信サービスの特征については、人の介入の有無、端末数、通信量、端末あたり収入、端末の移動範囲、通信のタイミング、そして、主な用途において、次のように異なっている。

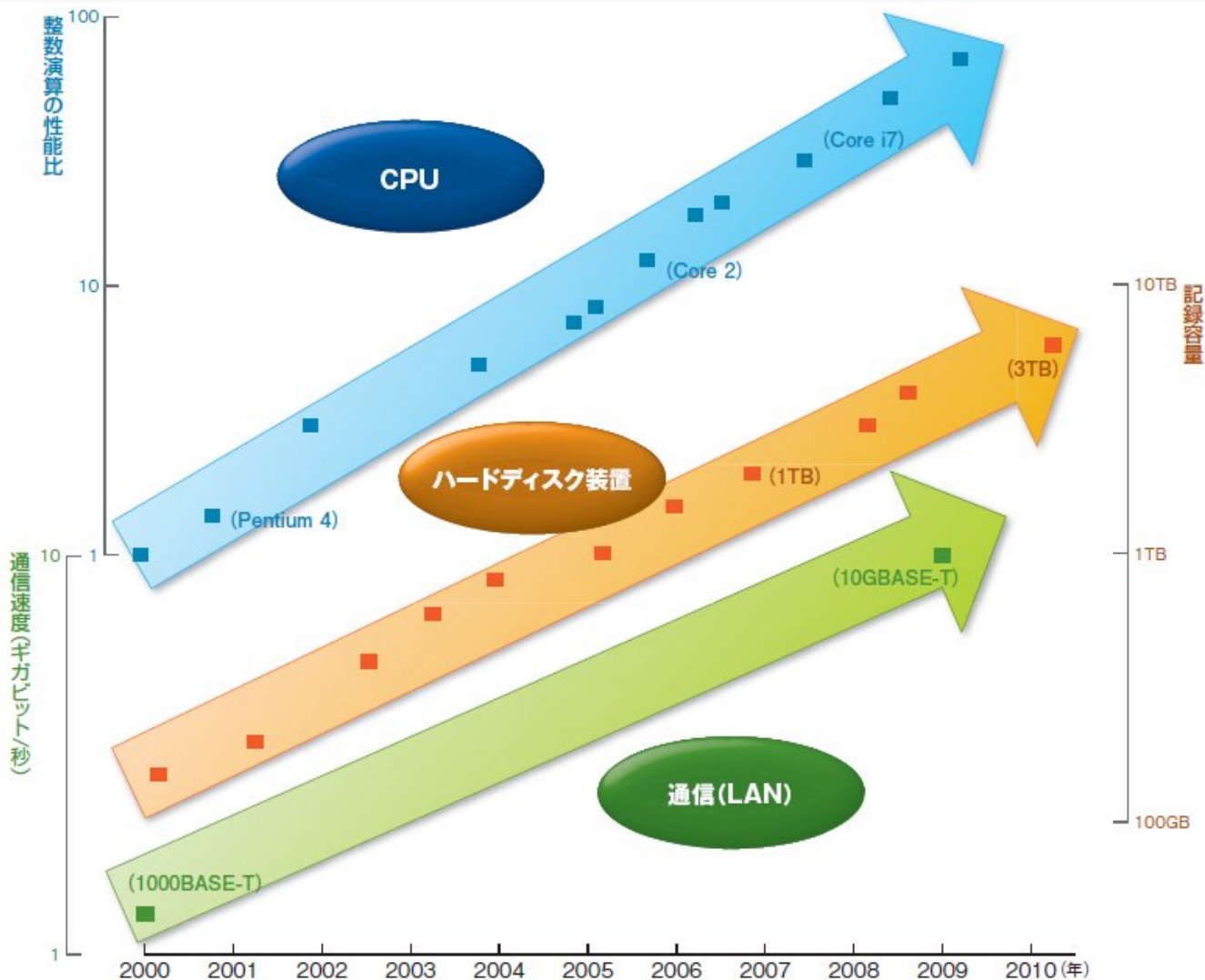
従来の通信サービスとM2M通信サービスの特征の違い

	従来の通信サービス(人)	M2M通信サービス(モノ)
人の介入	あり	なし
端末数	一定の上限が想定される	将来的には膨大な数が想定される (あらゆるモノ)
通信量	様々(音声、メール、音楽、動画) ~(上り)数Mbps程度/(下り)数十Mbps程度 <small>(電気通信事業者へのヒアリングより)</small>	極めて少ない ~(上り)数十kbps程度/(下り)数百kbps程度 <small>(電気通信事業者へのヒアリングより)</small>
端末あたりの収入 (ARPU*)	比較的高い(5,000円程度/月) <small>(電気通信事業者のウェブサイトより)</small>	低い(数百円程度/月) <small>(電気通信事業者へのヒアリングより)</small>
端末の移動範囲	非限定的(動き回る)	限定的(動かないことが多い)
通信のタイミング	ランダム (ユーザーニーズに合わせて随時)	集中的に発信される傾向
主な用途	スマホ、タブレット、デジタルフォトフレームなど	自販機、建設機械、エレベーターなどに付けられた 通信モジュール付きのセンサなど

* Average Revenue Per User: 通信事業者の1契約あたりの売上

ICTに関する性能の向上

● 2000年からの約10年間に於いて、CPUの整数演算の性能が約100倍、ハードディスク装置の記録容量が約100倍等、ICTに関するハードの性能が向上。



注：CPUの性能比は米インテルのPC向け製品の計測結果を、ハードディスク装置は3.5インチ型1台の記録容量を用いた

【出典】玄 忠雄、森山 徹「あなたの手にも新型を」（日経コンピュータ平成23年7月7日号）

ソーシャルメディアデータ

☞ ソーシャルメディアにおいて参加者が書き込むプロフィール、コメント等



マルチメディアデータ

☞ ウェブ上の配信サイト等において提供等される音声、動画等



ウェブサイトデータ

☞ ECサイトやブログ等において蓄積等される購入履歴、ブログエントリー等



カスタマーデータ

☞ CRMシステムにおいて管理等されるDM等販促データ、会員カードデータ等



ビッグデータ

ICT (情報通信技術) の進展により生成・収集・蓄積等が可能・容易になる多種多量のデータ (ビッグデータ) を活用することにより、異変の察知や近未来の予測等を通じ、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。

センサーデータ

☞ GPS、ICカードやRFID等において検知等される位置、乗車履歴、温度、加速度等



オフィスデータ

☞ オフィスのPC等において作成等されるオフィス文書、Eメール等



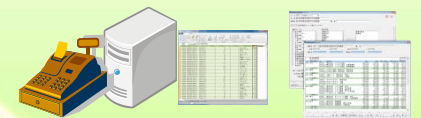
ログデータ

☞ ウェブサーバ等において自動的に生成等されるアクセスログ、エラーログ等



オペレーションデータ

☞ 販売管理等の業務システムにおいて生成等されるPOSデータ、取引明細データ等



ビッグデータの質的側面

- ビッグデータの特徴については、データの利用者やそれを支援する者それぞれにおける観点から異なるが、主に、多種性、リアルタイム性、多量性等が共通した特徴。

データを利用する者の視点から捉えた特徴(例)

[参考] 鈴木構成員説明資料(第4回)

- ☞ 個別に、即時に、多面的な検討を踏まえた付加価値提供を行うという観点において、事業に役立つ有用な知見として活用されるデータの特徴は、例えば、次の3つのとおり。

高解像

事象を構成する個々の要素に分解し、把握・対応することを可能とするデータ

高頻度

リアルタイムデータ等、取得・生成頻度の時間的な解像度が高いデータ

多様性

各種センサからのデータ等、非構造なものも含む多種多様なデータ

- ☞ 以上により、結果として、ビッグ（大きな）ボリューム（量）のデータが必要（**多量性**）

データの利用を支援する者の視点から捉えた特徴(例)

[参考] 日本アイ・ビー・エム説明資料(第2回)
エヌ・ティ・ティ・データ説明資料(第4回)
インテル説明資料(第5回)

- ☞ 以上のようなデータの利用者を支援するサービスの提供を行うという観点において、同サービスが対応可能なデータの特徴は、例えば、以上の**多量性**も含み、次のとおり。

多源性

複数のデータソースにも対応可能

高速度

ストリーミング処理が低いレイテンシーで対応可能

多種別

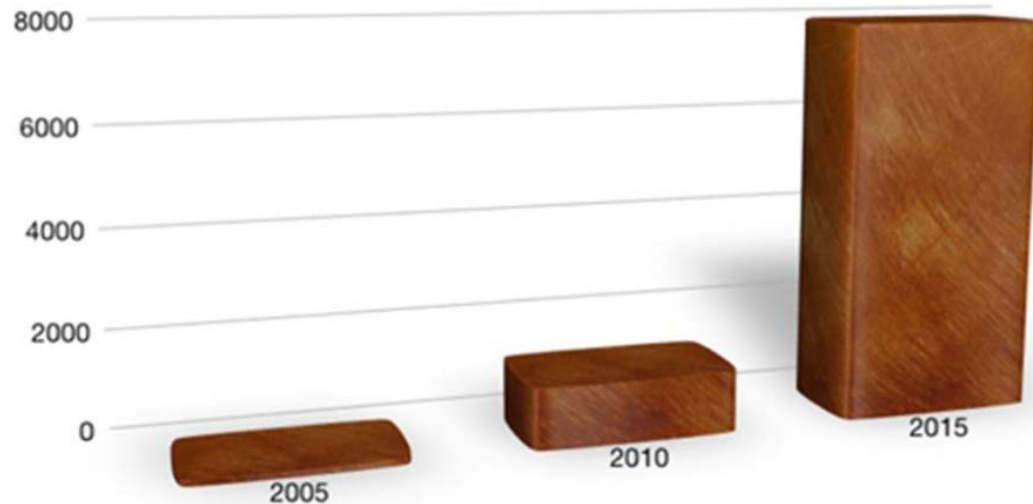
構造化データに加え、非構造化データも対応可能

ビッグデータの量的側面

- 国際的なデジタルデータの量は、2011年の約2ゼタバイト(2兆ギガバイト = 2千エクサバイト)から約4倍増加し、2016年には約8ゼタバイトへ拡大する見込み。
- ビッグデータの活用により、例えば、米国ヘルスケアで年間3千億ドル、EU公共セクターで年間2.5千億ユーロ、位置情報データの活用により年間6千億ドルの消費者価値創出等が期待。

10年間のデジタルデータの成長

(Exabytes)



いわゆる「ビッグデータ」の定量的価値(例)

50億台の携帯電話が使用(2010年)

300億のコンテンツが毎月Facebook上で共有

IT費用の5%増加で、年間40%増のデータ創出

米国のヘルスケアでは年間3000億ドルの価値創出が期待(スペインの年間ヘルスケアコストの2倍)

EUの公共セクターでは年間2500億ユーロの価値創出が期待(ギリシアのGDPを超える)

個人の位置情報データを活用することで年間6000億ドルの消費者価値創出が期待

小売の営業利益に60%改善の見込み

【出典：IDC「2011 Digital Universe Study: Extracting Value from Chaos」(平成23年6月)】

【出典：McKinsey Global Institute「Big data: The next frontier for innovation, competition, and productivity」(平成23年5月)】

NoSQL/Not only SQL

- ▶ 表形式によるリレーショナルデータベース管理システム(RDBMS)とは異なる設計によって実装されたデータベースシステム
- ▶ 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等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

Hadoop

- ▶ 米国NPOのApacheソフトウェア財団のプロジェクトで開発が進められている、大規模データの効率的な分散処理等のためのオープンソースソフトウェアフレームワーク
- ▶ 複数のサーバを通じた並列処理により、柔軟かつ継続的な大規模データの高速処理が可能



クラウドサービス

- ▶ 利用者が必要なコンピュータ資源を「必要な時に、必要な量だけ」利用でき、拡張性、可用性、俊敏性及び経済性等の特徴を有するサービス
- ▶ クラウドサービスの利用により、多種多量のデータの蓄積や計算処理のために必要となる多数のマシンについて、自前で用意する必要がなく、低コストで同様の環境の構築が可能
- ▶ また、例えば、「Hadoop」稼働環境を提供するサービスを利用すれば、マシンの調達に加え、ソフトウェアのインストール等の設定作業の省略も可能

DWH (Data Ware House)

- ▶ 定型データ・非定型データを問わず、大量データの蓄積を目的とするデータベースの総称
- ▶ これらの大量データを高速に処理する方法により、
 - ①従来のRDBMSとは異なる設計技術によるNoSQLデータベースと、
 - ②標準的なRDBMSとハードウェアレベルでの高速化技術を組み合わせたDWHアプライアンスの2つに大別

CEP (Complex Event Processing)

- ▶ データをディスクに格納せずに、書込速度がディスクに比べて高速なメモリ上で逐次的に処理することにより、必要な情報をリアルタイムに抽出する技術
- ▶ ディスクにデータを蓄積して分析する手法と比べ、短時間で処理が可能であるため、クレジットカードの不正利用や防犯カメラ映像の異常検知等、短期間での対応が必要となる場合に利用
- ▶ 予め利用者が定義するリアルタイム処理の内容には、単一のデータ属性の閾値による判別に加え、複数の属性を組み合わせた処理の設定も可能

PPDM(Privacy Preserving Data Mining)

- ▶ プライバシーを保護した上で、大規模なデータから特徴や規則性等を抽出する技術
- ▶ 匿名化や秘密計算等により、個々のデータを暗号化したままデータマイニングを実施することで、個人情報流出等のリスクを回避した上で、データの解析等が可能
- ▶ 例えば、どのデータについても同じものがk件以上存在するようにデータの粒度や曖昧さを制御するk-匿名化技術等が研究

MDM(Master Data Management)

- ▶ 業務を遂行する際の基本情報である顧客情報等のマスターデータを管理するためのシステム
- ▶ 多種多様な情報システムに重複・散在し、かつ、多量に生成等されるマスターデータについて、常時の最新状態への更新やシステム全体の整合性の確保等が可能

秘密計算

- ▶ 入力データや演算ロジックを暗号化したままで任意の計算を可能にする技術
- ▶ 複数のコンピュータにデータの断片を送り、断片の部分計算を繰り返し行うことで、データを秘匿したまま統計などの各種計算を行うことが可能
- ▶ PPDMでの利用の他に、民間企業、公共機関、教育現場などにおけるプログラムの不正解析防止、知的財産の侵害防止、情報漏えい防止などの多様な分野への応用が期待

アメリカ政府により、ビッグデータ活用に向けて、2億ドル以上の研究開発投資

～"Big Data Research and Development Initiative"を2012年3月29日に公表～

科学技術政策局(OSTP)が本イニシアチブを作成、下記をはじめとする多くの機関がこの取組をサポート。

- ▶ 大容量のデジタルデータの収集、保存、蓄積、管理、分析及び共有のためのツールと技術の向上を図る。
- ▶ ビッグデータを利用して、理工学の研究の加速、安全保障の強化、教育及び学習の改革の実現を目指す。

国立科学財団(NSF)



- ▶ データサイエンティスト育成のための大学でのプログラム。
- ▶ カリフォルニア大学での1,000万ドルのプロジェクト。
- ▶ 「Earth Cube」※プロジェクト支援のための助成金。

(※)地球科学者が地球に関する情報を利用、分析、共有できるシステム

エネルギー省



- ▶ 2,500万ドルかけて、「SDAV」(Scalable Data Management, Analysis and Visualization)という新たな研究機関を設立。

国立衛生研究所(NIH)



- ▶ 200テラバイトにもものぼる1,000のゲノムプロジェクトのデータがクラウド上で自由に手に入る。

国防高等研究計画局(DARPA)



- ▶ 年間2,500万ドルを4年かけて投資する「XDATA program」を実施し、非構造データを含めたデータ解析プログラムツールを開発。

国防総省



- ▶ ビッグデータ新規研究プロジェクトに6,000万ドル投資。
- ▶ ビッグデータに関するコンペティションを開催。

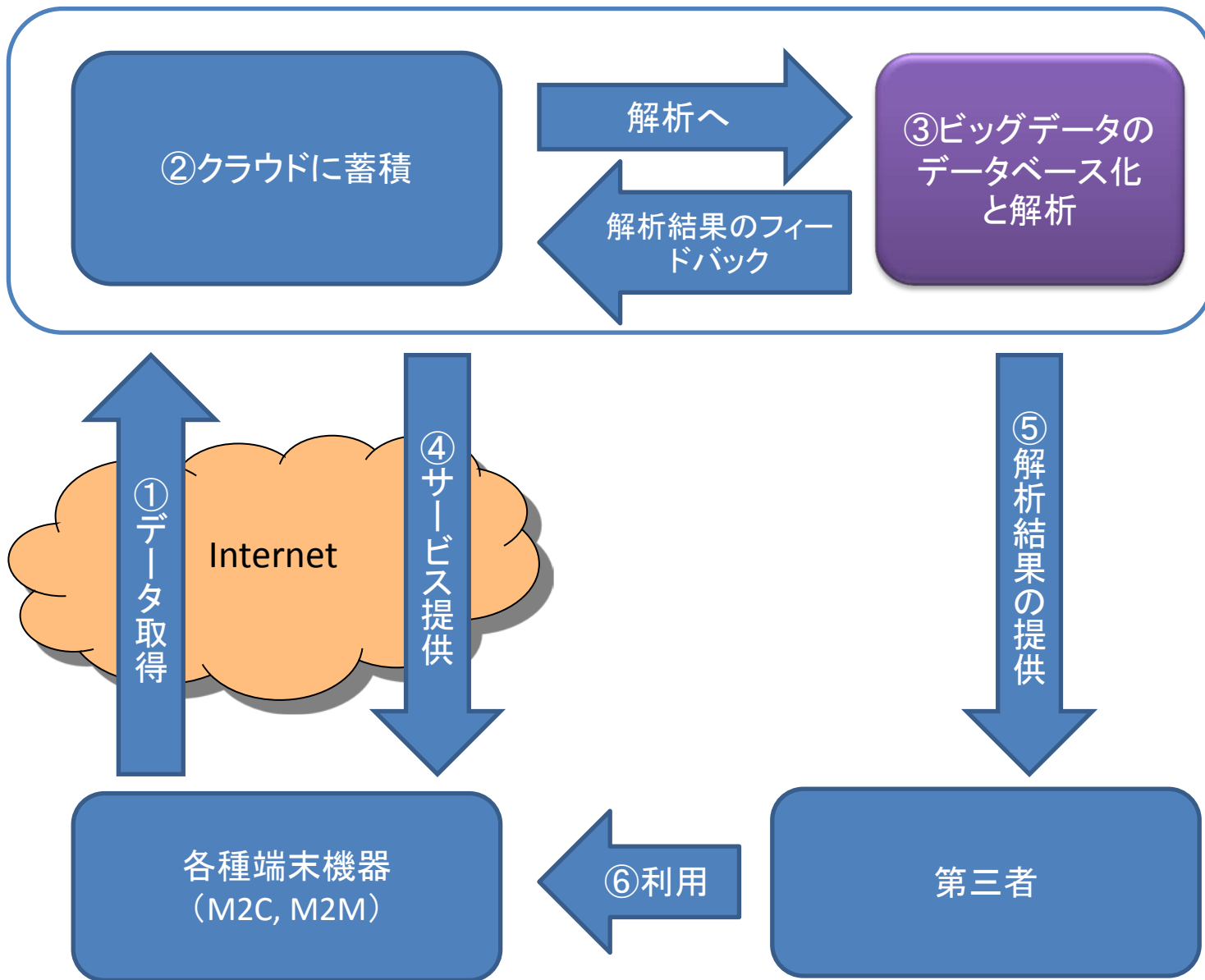
地質調査所



- ▶ 「John Wesley Powell Center」を提供することで、地球システム科学に関するビッグデータを分析等できる場を科学者に提供。

政府機関に加え、企業や大学、非営利組織にも、参加を呼びかけ

【想定モデル】



取り扱いに際して遵守すべき主要な法令(対ユーザ)

ユーザに対する関係で遵守すべき法令は、主として次の2グループ。

- ①個人情報保護法制、プライバシー権、通信の秘密、営業秘密。
- ②取得すべき個々のデータに関する著作権法上の権利。

M2Mの場合には、両グループともに原則として関係せず。

ビッグデータの法的保護(対第三者)

成果物が第三者によって流用された場合に、事業者は保護を受けられるか。

成果物は、ビッグデータをデータベース化したもの(中間成果物?)と、その解析結果(最終成果物?)とに分けて検討する必要。

前者については、創作性があれば、著作権法上のデータベース著作物として保護される場合あり。これに対し、創作性のないデータベース(ファクトデータベース)は、日本では同法による保護範囲外。

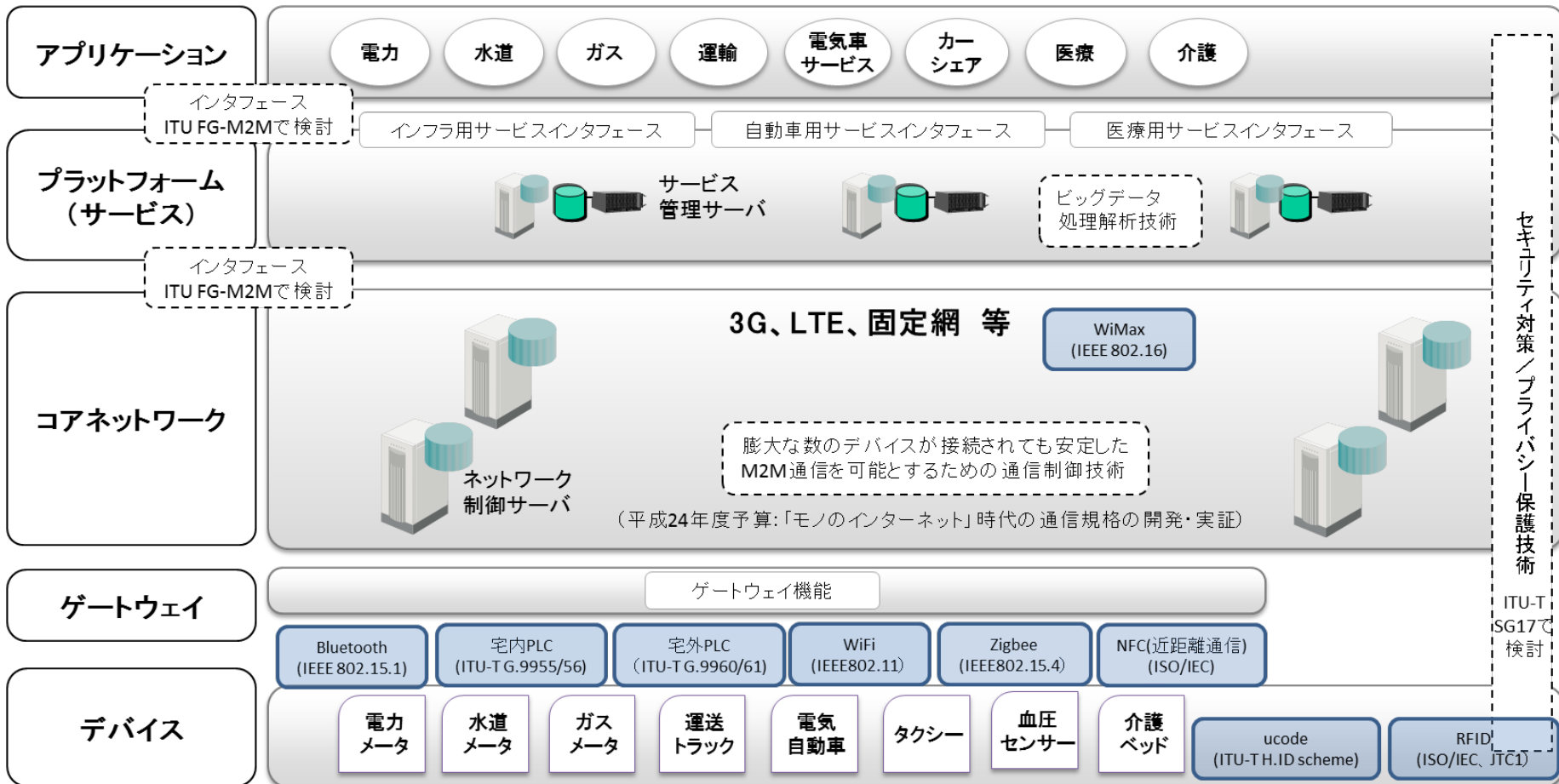
ビッグデータ解析結果の提供と法的責任(対提供先)

ベンダが他の者に提供した分析情報の内容が不正確だった場合の責任。

通常は責任減免条項によって対応。しかし、当該条項が有効となるとは限らない。

M2M通信サービスのアーキテクチャと技術的課題

- M2M通信サービスは、アプリケーション、プラットフォーム(サービス)、コアネットワーク、ゲートウェイ／デバイス等の各レイヤが相互に連携し、多種多様で膨大な情報の受け渡しを実現しており、各レイヤ間のインターフェースの標準化が必要。
- M2M通信サービスの普及・促進のためには、膨大な数のデバイスが接続しても安定的かつ安心・安全に運用が可能なネットワークの実現が技術的課題。



(凡例) 関連する通信規格 (Technical challenge)

ビッグデータの活用に関する人材

- 統計学や機械学習に関する高等訓練の経験を有し、データ分析を行うという深い分析に係る才能を有する大学卒業生数(2008年)について、国際的には、米国の2万4,730人、中国の1万7,410人、インドの1万3,270人に比べて、日本は3,400人。

新たに蓄積されたデータ量 (地域別)

New data stored¹ by geography, 2010
Petabytes

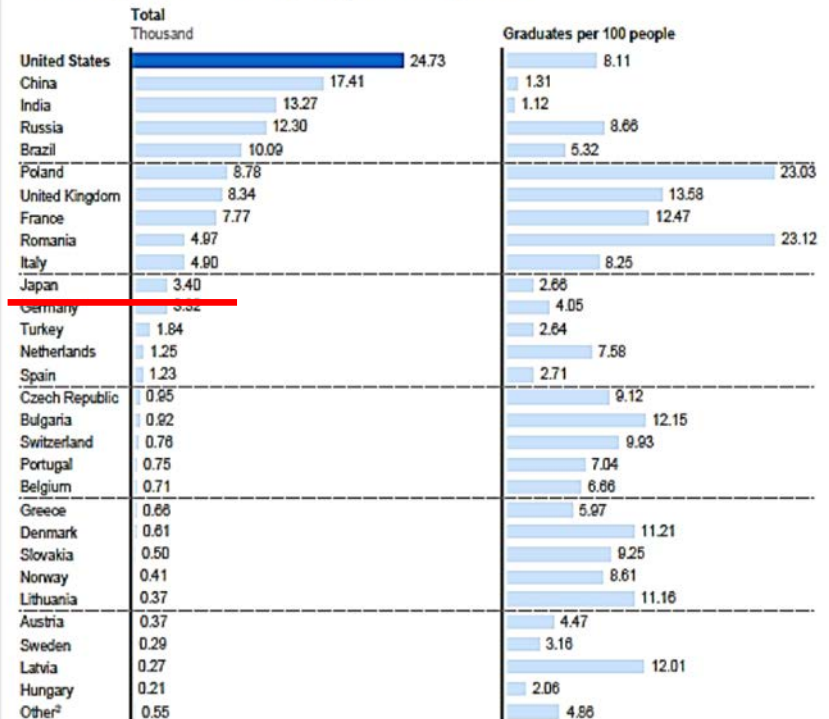


¹ New data stored defined as the amount of available storage used in a given year; see appendix for more on the definition and assumptions.

SOURCE: IDC storage reports; McKinsey Global Institute analysis

深い分析の訓練を受けた新たな大学卒業生数 (地域別)

Number of graduates with deep analytical training in 2008¹



¹ These data count new graduates, i.e., a flow of deep analytical talent, which we define as people with advanced training in statistics and/or machine learning and who conduct data analysis.

² Other includes Finland, Estonia, Croatia, Slovenia, Iceland, Cyprus, Macedonia, and Malta.

SOURCE: Eurostat; Russia Statistics; Japan Ministry of Education; India Sat; NASSCOM Strategic Review 2005; China Statistical Yearbook; China Education News; IMF World Economic Outlook Database

【出典：McKinsey Global Institute「Big data: The next frontier for innovation, competition, and productivity」(平成23年5月)】

2. ビッグデータの活用に関する国内外の取組と課題

データを把握する・
収集する

データを蓄積する処理
可能な状態にする

データを処理・分析する

活用



各種センサ技術の成熟



ミドルウェアとして通
信の活用およびクラ
ウドとの連携を促す
Android



移動体通信
モジュールの
活用



情報発信や閲覧・購
買・コミュニケーション
動態把握。(公開API
からのデータ収集)



二次利用可能な
統計データ等



人手による入力データや
クリックストリーム



業務付随データ
(事業者保有データ)

▼大量処理の基盤

クラウドコンピューティングサービス環境 (IaaSほか)

▼大量に蓄積する

DWH
データウェアハウス

▼データを分析可能に

MDM
マスターデータ管理

▼大量のデータを処理する

NoSQL /
 hadoop



Scipy NumPy...

トレーニングを重ねた
優れたアルゴリズム

▼秘匿と活用の両立

PPDM プライバシ保護データマイ
ニング(匿名化、秘密計算、再構
築計算)SDC 統計的開示抑制

▼たまずに、処理する(リアルタイム処理)

CEP(複合イベント処理)、ストリーム・コンピューティング

可視化技術

遠隔監視

需要管理

運行管理

決済管理

情報生成

.....

● 国内におけるビッグデータの活用に関する取組事例、課題や今後の方向性等について、自動車メーカー、損害保険会社、気象情報サービス事業者、電気通信事業者及びICTサービス事業者等の関係事業者10グループ(11者)よりヒアリングを実施。【ヒアリングの概要については、別添参照】

HONDA
The Power of Dreams

自動車の走行データの道路行政や災害対応への活用等について

Designing The Future
KDDI

トラフィック情報を基にした通信品質改善や海外における災害対応への活用等について

東京海上日動
docomo

携帯電話のGPSデータの保険サービスへの活用等について

weathernews

気象観測データや会員からのコメントデータの天気予報や災害対応への活用等について

NTT docomo

秘密計算による臨床研究データや、携帯電話ネットワークの運用データ等のまちづくり等への活用等について

FUJITSU

農地内センサーからの温度データ等の栽培管理や、健康関連データの予防医療への活用等について

IBM

センサー等からの大量なデータの公共インフラ管理や環境監視への活用等について

Preferred Infrastructure

リアルタイム性、大規模化及び深い分析に対応した処理基盤等による活用等について

NTT DATA

M2Mシステムや、橋梁の多数のセンサーによる歪等データの異常検知等への活用等について

intel

ビッグデータの活用に関する対応課題、インフラ基盤の在り方等について

- これまでの国内における取組事例に関するヒアリング等を踏まえると、ビッグデータの活用については、例えば、次の課題があると考えられる。

データの生成・収集・蓄積における課題(例)

- ☞ 同種のデータが生成されるセンサー等の管理者が様々である場合におけるデータ収集の在り方。
- ☞ 多種多様なセンサー等からのデータ収集等におけるセンサーの設計や統計的手法との連携等、フィジカルな領域におけるICT人材や統計等の専門家の活用の在り方。
- ☞ データの収集におけるGPSの普及やデータ入力に関するユーザインタフェース等の利用者からのデータ収集の在り方。
- ☞ 個人に関する情報を利用するサービスの提供における契約約款による同意の取得や重要事項の説明等の在り方。
- ☞ 例えば、スマートフォンにおけるユーザーからのパーミッションの取得時の説明等、通信の秘密や個人情報保護法上の「個人情報」に該当する場合における事前の利用目的の特定(同法第15条)、本人に対する利用目的の明示(同法第18条)や不正な手段による取得の禁止(同法第17条)への対応の在り方。
- ☞ センサー等で生成されるデータをリアルタイムに収集・処理するための帯域や遅延時間等のネットワークや処理システム等の基盤の在り方。
- ☞ センサー等からのデータ収集におけるネットワーク化するモノの多さとの費用対効果や、センサーをばらまくことによる100メートル以下における無線の競合の回避の在り方。
- ☞ 正確に特定された時間に関するデータの活用が必要とされる場合における利用者による確認の必要性等のネットワーク輻輳への対応の在り方。
- ☞ 企業等利用者におけるデータ活用の重要性の認識やIDによる紐付け等のデータ整備・管理の在り方。
- ☞ 複数主体からの多種多量のデータをクラウド等により取扱う場合における適切なデータ管理の在り方。

データの流通・連携における課題(例)

- ☞ 情報を的確に常時送る場合における放送メディア等の全国的なインフラ整備の在り方。
- ☞ 指数関数的に増加するデータトラフィック等ビッグデータを支える基盤となるサーバ、ストレージ及びネットワーク等について、ソフトウェアでフローをコントロールする技術やHadoop等の各種ストレージ技術等の活用や、電力等を考慮したアーキテクチャ等の在り方。
- ☞ 携帯電話ネットワークを活用したデータの送受信やサービス販売のプラットフォームとしての携帯電話の活用におけるスマートフォンへの対応の在り方。
- ☞ 例えば、契約約款の変更について事前同意がある場合における当該約款の内容による利用目的の変更や統計的利用のための完全匿名化等、個人情報保護法上の「個人情報」に該当する場合における第3者への提供時の事前の同意取得(同法第23条)への対応の在り方。
- ☞ 位置情報等の情報自体やそれに基づく一定の事実のみの提供等、情報の種類に応じたリスクとビジネスチャンスを踏まえた、情報の複数事業者間における授受等の取扱いの在り方。
- ☞ 例えば、ツイッターの内容の第3者への表示等、第3者に提供すべきデータの解析結果にユーザから取得した著作物が含まれている場合における表現上の本質的特徴が直接感得できないものに情報提供の内容を変更することや約款によりユーザから許諾を取得すること等の対応の在り方。
- ☞ 震災時等において、自動的に、様々な者からデータがアップされサーバ間で連携・加工されることにより、必要な情報が関係者に送信される公共的なプラットフォームや体制等の在り方。
- ☞ 民間企業において収集等されたデータを平時から公開等により流通・連携する場合におけるデータ形式の標準化やメタデータの定義化、先行者利益に配慮したビジネスモデルの在り方。
- ☞ 国民IDの将来的な民間利用等のIDによる個人に関する情報の紐付けの在り方。
- ☞ 各分野におけるルールや情報リテラシー等の相違等を踏まえた情報共有のための入力や表示等の在り方。

その他のデータの活用における課題(例)

- ➡ 海外事例や海外企業による日本国内でのサービス提供状況を踏まえた、利用者との明示的な契約の締結や統計的な匿名化処理等による、災害対応等の公共目的のための通信ログの利用の在り方。
- ➡ 通信サービスの提供にあたっての位置登録情報等の運用データの非識別化処理等の技術やまちづくり等の公共分野における活用の在り方。
- ➡ EUからのデータ移転や国内における医療分野等における機微情報の取扱い等、個人に関する情報の取扱いの在り方。
- ➡ EUのデータ保護指令の改正やアメリカにおけるDo Not Track等の全体的な国際動向を踏まえた対応の在り方。
- ➡ 例えば、プライバシーポリシーの統合によるサービス提供や児童からの同意取得の場合や統計的利用のための完全匿名化等、個人情報保護法上の「個人情報」に該当する場合における利用目的の範囲内での利用(同法第16条)への対応の在り方。
- ➡ 個人識別性がない情報が転々流通するうちに個人識別性を有する可能性やプライバシーを侵害する可能性があるライブログ活用サービス等の提供にあたっての透明性の確保や利用者関与の機会の提供等の配慮の在り方。
- ➡ データ内容の暗号化とIDの取扱いや、k-匿名化に関する技術的な研究等の個人に関するデータに関する秘匿化の在り方。
- ➡ 誤った情報が利用者へ送信される場合等における情報の受信端末等の更新や、契約における責任減免条項の規定等、利用者における判断と提供側の責任等の責任分界の在り方。
- ➡ データの正確性や観測器の精度等が制度上求められている気象データ等をインターネット上で公開等することにより他分野や他目的で活用する場合におけるデータの正確性の確保等の在り方。

その他のデータの活用における課題(例)[続き]

- ➡ リアルなサービスにおける各種ログの収集や、多様なデータを組み合わせて活用できる仕組みの在り方。
- ➡ 災害関連や人の流れ等の多様な情報について、個人情報保護やプライバシー、著作権に配慮しつつ、まずは情報を使えるようにするための提供フォーマットやインターフェース等、政府や地方自治体が保有するデータの一元的な公開の在り方。
- ➡ 公共データにおけるデータの所在の明確化やデータ取得時の形式等のプロトコルの在り方。
- ➡ Hadoop等のオープンソースソフトウェアに関する運用コミュニティが買収された場合やサポート主体の不在への対応の在り方。
- ➡ ホストコンピュータにあげる前処理におけるメタデータ化に関するフォーマット変換等、今後のよりリアルタイムなデータ処理等における高度かつ新たな処理基盤と分析技術の在り方。
- ➡ 現在の利用目的毎かつ様々なデータ形式によるアプリケーションセントリックなデータの活用とは異なり、今後のM2M等の様々なデータが活用される場合において個々のアプリケーションのベースとなるセキュリティやプライバシー等の水平部分の標準化等の在り方
- ➡ 収集等したデータの帰属先や、当該データに関する第3者によるフリーライドや流用への対応の在り方。
- ➡ 多大な投下資本が必要となる一方で、情報の選択又は体系的な構成による創作性を有しないために著作物として保護されず、また、損害賠償請求の対象にはなる場合があるが、差止請求の対象にはならないファクトデータベースに関する欧米等の諸外国の状況も踏まえた保護の在り方。
- ➡ 営業秘密としての管理と、守秘義務契約による個別企業への提供や第3者への提供の関係等、著作物として保護されないデータ解析結果に関する取扱いの在り方。
- ➡ 日本の大学に統計学部や学科が存在せず、工学部・経済学部・医学部等の各学部で統計関連のスタッフが在籍し、横のつながりが少ない状況等における非構造化データ等の様々なデータの活用に関する数学的・統計学的・法学的な知識やビジネス管理における知識等を有する人材等の確保・育成の在り方。

その他のデータの活用における課題(例)[続き]

- ☞ 行政機関において民間から購入する等活用し得るデータの在り方。
- ☞ データの誤用・不適切利用におけるアルゴリズムの賢さ・間違いの取扱いの在り方。
- ☞ 量や多様性の中から価値ある情報を見つけるための分析ツール、データ分析のための継続的なノウハウの蓄積、そして、適切なデータの取得のための工夫を実施するためのビジネス等への総合的な理解と深い分析技術による仮説構築・検証・データ取得方法等に関するPDCAサイクルの継続的な検証の在り方。
- ☞ 国内におけるスモールサクセス等の成功体験や海外事例の共有等、データ活用の重要性に関する意識向上の在り方。
- ☞ 自社内で膨大にたまっているデータを新規事業に役立てることができる確率を高めるための他業種等との連携の在り方。
- ☞ 例えば、社会インフラや医療分野等、新産業の創出やビジネス領域の拡大をもたらす分野への投資を促進することによるマーケットセグメントを跨がるデータの融合を促すための仕組みの在り方。
- ☞ フィジカルとの融合等、GoogleやFacebook等が現在主に取り組んでいない分野への取組の在り方。
- ☞ 個人情報の活用におけるベネフィットと失われるものとのバランスの問題に関する実証の場等、国民の不安感の払拭やコンセンサス等の醸成の在り方。
- ☞ 実社会において実際に収集したデータを用いた分析結果や実証事業をアクション等するためのオープンイノベーションに向けた場の在り方。
- ☞ 海外プロジェクトにおける技術実証の進展も踏まえた、国内における技術の実装のためのソーシャルな利益をもたらすビッグデータの活用に関する社会実験等の在り方。
- ☞ スマートフォンの普及やテロ対策としてのソーシャルモニタリング等により増加するウェブにおける多言語化への統計的自然言語処理等による多言語解析の在り方。

3. ビッグデータの活用に向けた方向性と具体的方策

ビッグデータの活用に向けた今後の方向性

- 関係事業者からのヒアリング等による国内外の取組事例について、①データ分析結果のフィードバックまでの期間（ストック型/フロー型）、②データ分析結果のフィードバック対象(系全体/個別)で、便宜的に整理すると次のとおり。
- 今後、ビッグデータの活用については、系横断的なデータの活用への進化、活用のよりリアルタイム化への進展に伴い、社会的課題の解決や経済規模の拡大に貢献していくと考えられる。

系横断的なデータの活用への進化

活用のよりリアルタイム化の進展

	蓄積したデータの分析結果をフィードバック 【ストック型】	データのリアルタイムな分析結果をフィードバック 【フロー型】
データ分析結果 系全体への フィードバック	<ul style="list-style-type: none"> ● 減災レポート【ウェザーニューズ】 ● 図書名寄せ【PFI】 ● 電子書籍ハイライト情報【Amazon】 ● 陳列棚配置の最適化【shopperception】 	<ul style="list-style-type: none"> ● インターナビ【本田技研工業】 ● Smarter Planet【日本IBM】 ● ソーシャルマーケティング【PFI】 ● モバイル空間統計【NTTドコモ】 ● ロードコントロールシステム【ANA】 ● 東京ゲートブリッジ【国交省・東京都・NTTデータ】 ● 犯罪予測システム【サンタクルーズ市(アメリカ)】 ● Dynamic Discount Solution 【エリクソン・MTN(南アフリカ)】 ● 電気自動車の最適ルート案内【フォード(アメリカ)】
データ分析結果 個別への フィードバック	<ul style="list-style-type: none"> ● ECサイト分析・ウェブアクセス履歴 【PFI、リクルート、楽天】 ● ゲリラ雷雨予測【ウェザーニューズ】 ● 疾病予防管理サービス【徳島大学病院】 ● 契約者分析による解約防止 【Tモバイル(アメリカ)】 	<ul style="list-style-type: none"> ● ソラテナ【KDDI・ウェザーニューズ】 ● ドコモワンタイム保険 【NTTドコモ・東京海上日動火災保険】 ● みかん栽培【早和果樹園・富士通】 ● 投資情報SNS分析【カブドットコム証券・日本IBM】 ● 建機の稼働状況の遠隔監視【コマツ】 ● 新生児集中医療【オンタリオ工科大学(カナダ)】 ● クレジットカードの不正検知【VISA(アメリカ)】

(注) 「フロー型」における「リアルタイム」の範囲については、業種・業態、データの活用目的により異なるものであるため、実際は、「フロー型」と「ストック型」の2つに明確に分類されるわけではない。

● ビッグデータの活用に関する市場規模等の計測手法については、国際的に確立されていない状況であるが、諸外国に関する民間調査機関による試算等を前提とした場合の日本における効果として、データの利用事業者及びその支援事業者からなるビッグデータの活用に関する市場においては、今後、少なくとも10兆円規模の付加価値創出及び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る。



▷ 業務、事業への適用による業務効率化、付加価値創出

		我が国における発現効果(試算)	
※ 民間調査機関の 試算における 対象範囲	医療	・医療費最適化:	3.1～4.6兆円
	行政	・行政効率化: ・社会保障給付是正: ・租税増収:	7,200億円～1.2兆円 2,995.5億円～1.2兆円 2,133.9～8,535.6億円
	小売	・利益増加額:	0.95兆円以上
	製造	・製品開発費節減:	最大5.7兆円
	位置 情報	・サービス収入: ・エンドユーザー価値:	3,040億円 2.1兆円
その他の分野	交通	・プローブ交通情報の導入による渋滞解消効果 渋滞による経済損失(11兆円:H17年国交省) ×走行時間削減効果(19%:NRI「全力案内」)	2.09兆円

※McKinsey Global Institute 「Big data: The next frontier for innovation, competition, and productivity」 (平成23年5月)

- ▷ M2M(2020年に約9,000億円)
- ▷ ……
- ▷ クラウドサービス
(2016年に2.8兆円、単純外挿すると2020年までに4.2兆円)
- ▷ ストレージ関連ソフトウェア(2020年に約977億円)
- ▷ ……
- ▷ ビジネスインテリジェンスツール(2020年に1,940億円)
- ▷ ……

【出典：(株)野村総合研究所説明資料(第6回)より作成】

【ビッグデータの活用に関する現状と今後の方向性】

- ビッグデータの活用については、現在、検索、EC、ソーシャルメディア等のウェブサービス分野において多量に生成・収集等されるデータを各種サービスの提供のために活用することを中心に進展。
- 今後は、それらのデータや技術も活用しつつ、M2M等のセンサネットワーク等から生成・収集等される多種多様なデータを実社会分野において系横断的・リアルタイムに活用することが進展する見込み。

【ビッグデータの活用を推進することの必要性】

- 他方で、競争の激化等が進展する国際経済・社会において、人口減少等により今後国を支える人的資源が縮小し、また、東日本大震災を契機として情報が命を守るライフラインであることが再認識されている状況。
- ものづくりをはじめとする日本の強みを活かしつつ国際競争力を強化し、更なる成長を実現するためには、ビッグデータを戦略的な資源と位置づけ、個人情報等にも配慮しつつ、国としても実社会分野におけるビッグデータの活用を積極的に推進することが重要。

【ビッグデータの活用におけるICT政策の役割】

- その上で、ICT政策としては、国、地方自治体、公共・民間事業者等のそれぞれにおいてM2M等を通じ生成・収集等される多種多量のデータについて、社会全体で共有可能な知識や情報の創発が促進されるよう生成・収集・蓄積・公開・流通・連携等させることを通じ、社会的課題の解決や経済活性化の実現に貢献すべき。
- なお、以上にあたっては、昨今の個人に関するデータの取扱いを巡る問題等、実社会への適用や技術開発の進展状況等に関する国際的な動向も見極めつつ、制度的・技術的課題の解決等に取り組むことが必要。

● ビッグデータの活用における基本的な考え方を踏まえると、ICT政策としては、例えば、次のような7つの課題の解決に向けて取り組むことが必要。また、それら以外の課題については、引き続き民間分野における取組を注視することが必要。

- ① 多様な分野において閉じた形で保有されているデータについて、オープンガバメントの推進等官民におけるオープンデータ化、街づくりや防災等への活用等横断的活用のための環境整備の在り方
- ② リアルタイムで活用するビッグデータについて、センサ等から生成されるデータを安心・安全に収集・解析・流通等するための基盤技術の研究開発・標準化の在り方
- ③ 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ビッグデータの活用に関する人材について、産学官のプロジェクトを通じた育成等による確保の在り方
- ④ ビッグデータビジネスの創出に寄与するM2M（人が介在せず、ネットワークに繋がれた機器同士が相互に情報交換等を行う機器間通信）の普及促進の在り方
- ⑤ 正確性の確保等のために多様な用途への転用が制限されているデータや既存制度の保護対象とならないため整備が進まないデータ等について、その活用を阻む規制・制度の在り方
- ⑥ 様々な業種の民間事業者、研究機関、学識経験者、行政機関等から広く構成され、データ資源の蓄積等を通じて、ビッグデータの活用について国内の普及・展開を図るための推進体制の在り方
- ⑦ 国際的な取組事例等の共有等を図るための外国政府等との意見交換の在り方や、ビッグデータの活用による経済価値の見える化等のための計測手法の在り方

① 多様な分野において閉じた形で保有されているデータについて、オープンガバメントの推進等官民におけるオープンデータ化、街づくりや防災等への活用等横断的活用のための環境整備の在り方

具体的方策

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）

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

今後の推進に向けたアクション

- 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の開発及び国際標準化を推進。

- 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及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。

② リアルタイムで活用するビッグデータについて、センサ等から生成されるデータを安心・安全に収集・解析・流通等するための基盤技術の研究開発・標準化の在り方

具体的方策

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

今後の推進に向けたアクション

- 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。

- ③ 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ビッグデータの活用に関する人材について、産学官のプロジェクトを通じた育成等による確保の在り方

具体的方策

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

今後の推進に向けたアクション

- 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通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及び若手研究者やベンチャーへの開放。

- ④ ビッグデータビジネスの創出に寄与するM2M（人が介在せず、ネットワークに繋がれた機器同士が相互に情報交換等を行う機器間通信）の普及促進の在り方

具体的方策

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

今後の推進に向けたアクション

- 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。

- ⑤ 正確性の確保等のために多様な用途への転用が制限されているデータや既存制度の保護対象とならないため整備が進まないデータ等について、その活用を阻む規制・制度の在り方

具体的方策

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

今後の推進に向けたアクション

- 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、下記⑥の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。

- ⑥ 様々な業種の民間事業者、研究機関、学識経験者、行政機関等から広く構成され、データ資源の蓄積等を通じて、ビッグデータの活用について国内の普及・展開を図るための推進体制の在り方

具体的方策

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

今後の推進に向けたアクション

- 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。

- ⑦ 国際的な取組事例等の共有等を図るための外国政府等との意見交換の在り方や、ビッグデータの活用による経済価値の見える化等のための計測手法の在り方

具体的方策

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

今後の推進に向けたアクション

ビッグデータの活用に関する計測手法の確立

- 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、調査手法及び評価手法の確立