

ビッグデータの 社会活用推進上の課題に関する考察

2018年1月24日
中央大学研究開発機構
才所敏明 辻井重男

1

発表内容

- (1)ビッグデータ社会に向けて
- (2)ビッグデータシステムを構成する
サブシステムの現状・動向・課題
- (3)ビッグデータ活用推進上の今後の課題

2

(1)ビッグデータ社会に向けて

「日本再興戦略2016」(平成28年6月閣議決定)

“今後の生産性革命を主導する最大の鍵は、IoT(Internet of Things)、ビッグデータ、人工知能、ロボット・センサーの技術的ブレークスルーを活用する「第4次産業革命」である”

600兆円に向けた「官民戦略プロジェクト10」の一つとして、新たな有望成長市場の創出のための「**第4次産業革命(IoT・ビッグデータ・人工知能)**」の推進を掲げている。**<付加価値創出:30兆円(2020)>**

「世界最先端IT国家創造宣言・官民データ活用推進基本計画」

(平成29年5月閣議決定)

「官民データ利活用社会」～データがヒトを豊かにする社会～モデルを世界に先駆けて構築を目指し、「**我が国の置かれた諸状況を踏まえたデータ利活用による新たなライフスタイルの提案**」、「**官民データの利活用に向けた環境整備等の施策**」を掲げている。

「未来投資戦略2017」(平成29年6月閣議決定)

「未来投資戦略2017」においても、Society 5.0に向けた課題の一つとして、「**データ利活用基盤の構築、徹底したデータ利活用に向けた制度整備**」を掲げている。

(1)ビッグデータ社会に向けて

「第5期科学技術基本計画」(平成28年1月閣議決定)

“超スマート社会サービスプラットフォームの構築に必要となる基盤技術、すなわち**サイバー空間における情報の流通・処理・蓄積に関する技術**は、我が国が世界に先駆けて超スマート社会を形成し、**ビッグデータ等から付加価値を生み出していく上で不可欠な技術**である”

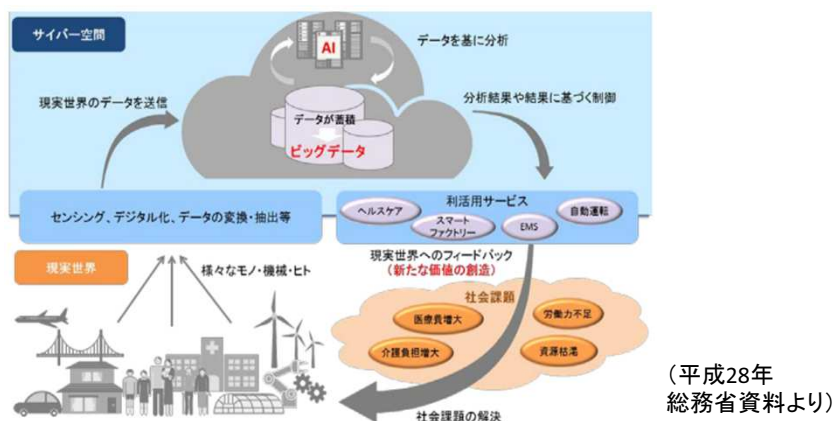
とし、非構造データを含む多種多様で大規模なデータから知識・価値を導出する「ビッグデータ解析技術」の強化を図る、としている。

(Society5.0の提唱:ICTを最大限に活用し、サイバー空間とフィジカル空間(現実世界)とを融合させた取組により、人々に豊かさをもたらす「超スマート社会」を未来社会の姿として共有し、その実現に向けた一連の取組を更に深化させつつ「Society 5.0」として強力に推進し、世界に先駆けて超スマート社会を実現していく。)

「科学技術イノベーション総合戦略2017」(平成29年6月閣議決定)

新たな経済社会である Society 5.0 を実現していくためには、新たな価値創出を容易とするプラットフォームを構築することが重要とし、取り組む必要がある施策として「**新たな価値やサービスの創出の基となるデータベースの構築と利活用**」、「**プラットフォームを支える基盤技術の強化**」などを掲げ、“特に、**AI技術、IoTシステム構築技術、ビッグデータ解析技術等のいわゆるAI関連技術は Society 5.0 を実現する鍵**であり、世界の先を見据えた水準に昇華させ、更に社会実装を迅速に推進することが肝要である”としている。

(1)ビッグデータ社会に向けて



ビッグデータは、従来の情報システムとは異なるデータソース(モバイルデバイスやIoT等)からのデータとの結合によるデータフュージョンの発生や高度な分析を可能とする人工知能技術(AI)の発展により、ビッグデータの活用はますます加速する傾向にある。ビッグデータ、IoT、AIは相互に連携しながら発展をつづけ、データ・ドリブン(駆動型・主導)社会、ビッグデータ社会へと発展することになる。

5

発表内容

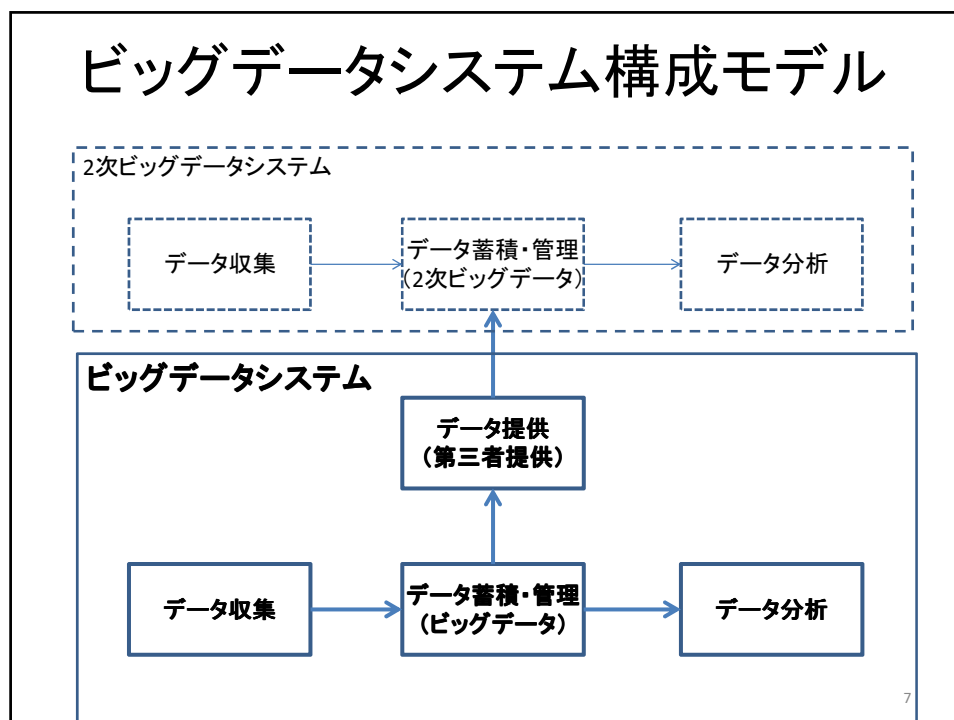
(1)ビッグデータ社会に向けて

**(2)ビッグデータシステムを構成する
サブシステムの現状・動向・課題**

(3)ビッグデータ活用推進上の今後の課題

6

ビッグデータシステム構成モデル



データ収集サブシステム

- (1) データ量の爆発
- (2) データ構造の多様化
- (3) データの品質確認・真正性確保
 - データの品質の悪さ、品質のばらつき
 - 誤データの混入、悪意のあるデータ注入
 - ⇒ 分析結果への悪影響
 - データ収集主体の責任問題
- (4) 個人情報・要配慮個人情報の適正取得
 - ⇒ 個人情報の取得: 利用目的を、
 - 本人に通知または公表することが必要
 - ⇒ 要配慮個人情報の取得: 本人の同意が必要
 - ＜改正個人情報保護法: 平成29年5月30日施行＞

データ蓄積・管理サブシステム

- (1) 多様な構造の大量なデータの蓄積・管理
- (2) データの維持・保全
 - データの破壊や改ざん、ランサムウェア
 - ⇒ アクセス者の認証と認可、バックアップ
- (3) データの漏洩防止
 - データへの不正アクセス
 - ⇒ 認証と認可、暗号化、秘密分散、匿名化
 - バックアップデータ ⇒ 暗号化、秘密分散
- (4) 保有個人データに対する本人の要求への対応
 - 開示、訂正、利用停止等

9

データ分析サブシステム

- (1) 統計学に基づく分析
 - (2) 人工知能・機械学習・深層学習による分析
 - 一般に分析結果の説明が困難
 - ⇒ 分析結果の導出根拠技術
 - (3) 分析過程のデータ保護(セキュア分析)
 - ⇒ 暗号化状態処理、秘密分散状態処理
 - (4) 分析結果の利用に関する責任の所在
 - ⇒ 分析結果の導出根拠技術
 - ⇒ 責任の所在に関する事前の合意
- (データ収集主体、分析機能提供主体、分析者、利用者)

データ提供(第三者提供)サブシステム

(1) 実名のまま提供

個人情報: 原則本人の同意による提供

オプトアウトによる提供も可

要配慮個人情報: 本人の同意による提供

(オプトアウトによる提供は認められていない)

(2) 匿名化し提供

匿名加工情報

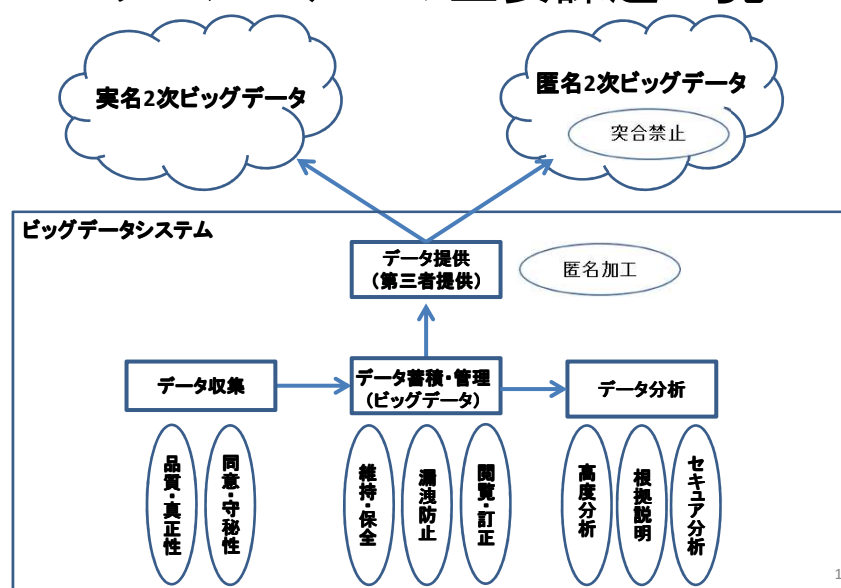
個人を識別できない/個人情報を復元できない

本人の同意は不要

匿名加工情報に含まれる情報項目の公表

11

ビッグデータシステムを構成するサブシステムの主要課題一覧



発表内容

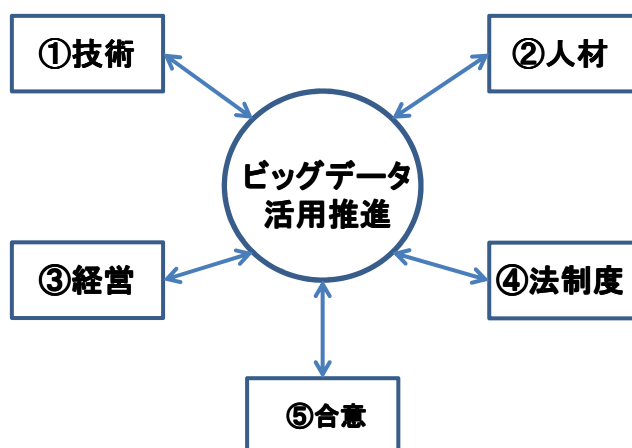
(1)ビッグデータ社会に向けて

(2)ビッグデータシステムを構成する
サブシステムの現状・動向・課題

(3)ビッグデータ活用推進上の今後の課題

13

ビッグデータ活用推進を支える基盤・環境



14

①技術

- (1)人工知能等による高度分析技術
および分析結果の根拠提示技術
- (2)暗号化状態処理等によるセキュア分析技術
- (3)個人データの
第三者提供のための匿名加工技術
- (4)集めないビッグデータに関する技術
個人のデータは個人の管理下へ(PDS)
情報銀行:個人のデータの管理・運用代行
- (5)ブロックチェーンに関する技術
BlockchainDB:分散型DB(MongoDB)に
ブロックチェーンの特徴的機能を実装
Keyless Signature Infrastructure(KSI):
ハッシュを利用してデータを繋ぐ
ブロックチェーン

①技術

BlockchainDB

開発元:BlockchainDB GmbH(ベルリン)

発表:2016年2月

開発目標:大企業が求める処理能力、分散型、
一度記録されたデータは不変、
アセットの登録や移転が可能

開発アプローチ:大量のデータを扱う分散型DBに
ブロックチェーンの特徴を実現

パートナー企業:世界で20社以上

(日本企業:リクルート、トヨタ)

16

①技術

Keyless Signature Infrastructure(KSI)

開発元: Guardtime (エストニア)

開発: 2007年

開発目標: 完全性、透明性、監査性を重視

ドキュメントのハッシュ値による

ドキュメントの非改ざん性の検証が可能

(ドキュメントの署名者の検証は別途)

開発事例: エストニアの100万人規模の

非改ざん性を検証可能な医療情報DB

開発アプローチ: OracleDBへのKSI add-onの開発

17

Bitcoin is harmful!

トランザクション1件処理に必要な電力消費量は、

VISAの5000倍以上が必要

トランザクション1件処理に必要な電力消費量は、

米国家庭9軒の1日分に相当

現在のビットコインネットワークの計算能力の総計は、

世界の高速スーパーコン500番目までの合計の10万倍以上

現在、ビットコインネットワークの電力消費量は31 terawatt-hours/year

1日当たり450gigawatt-hours増加中

(ハイチの全国の電力消費量の1年分に相当!)

現在のペースでビットコインネットワークが電力使用量が増加すると...

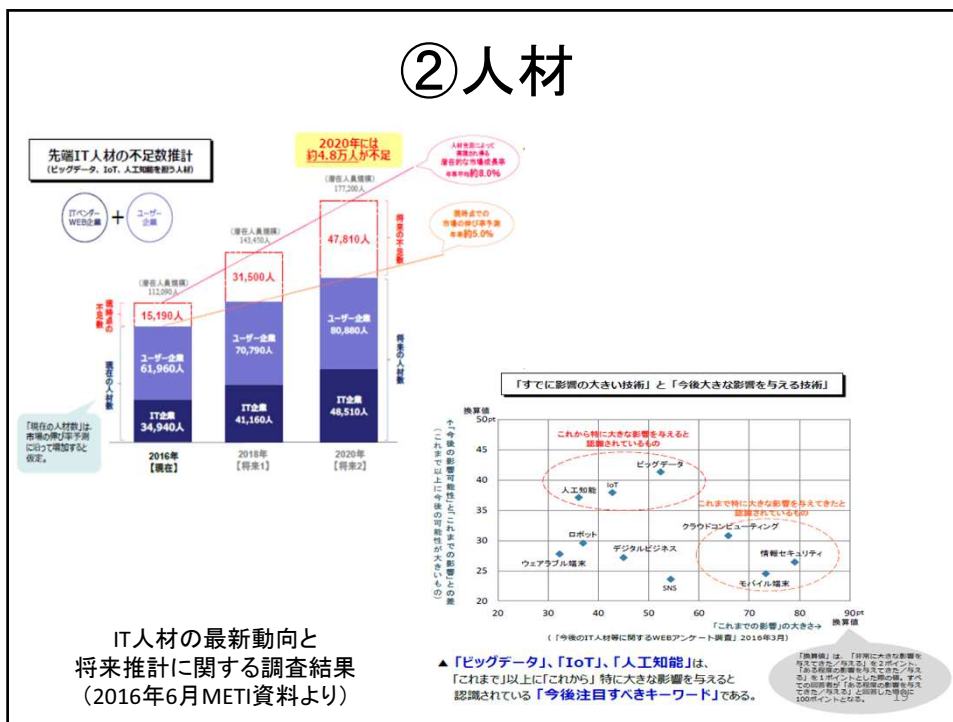
2019年7月までに、

米国全体の現在の電力を上回る電力を必要となる見込み

2020年2月までに、

世界全体の現在の電力と同程度の電力を使用することになる

②人材



IT人材の最新動向と将来推計に関する調査結果 (2016年6月METI資料より)

②人材

2012年のガートナー報告:

当時、日本には千人程度のデータサイエンティスト。将来、日本では25万人程度、不足する。

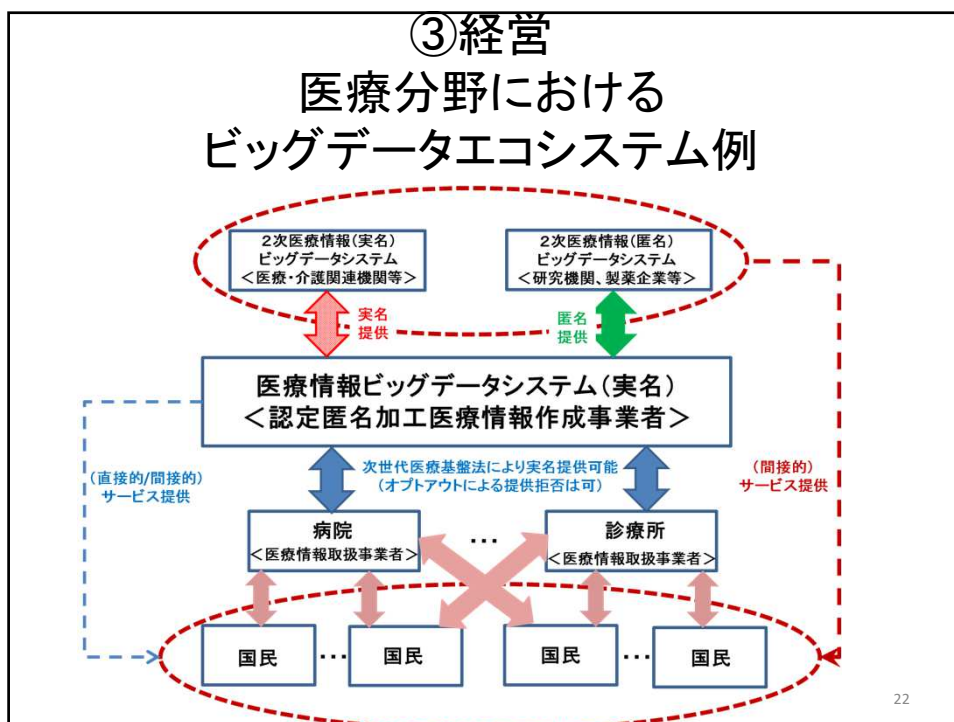
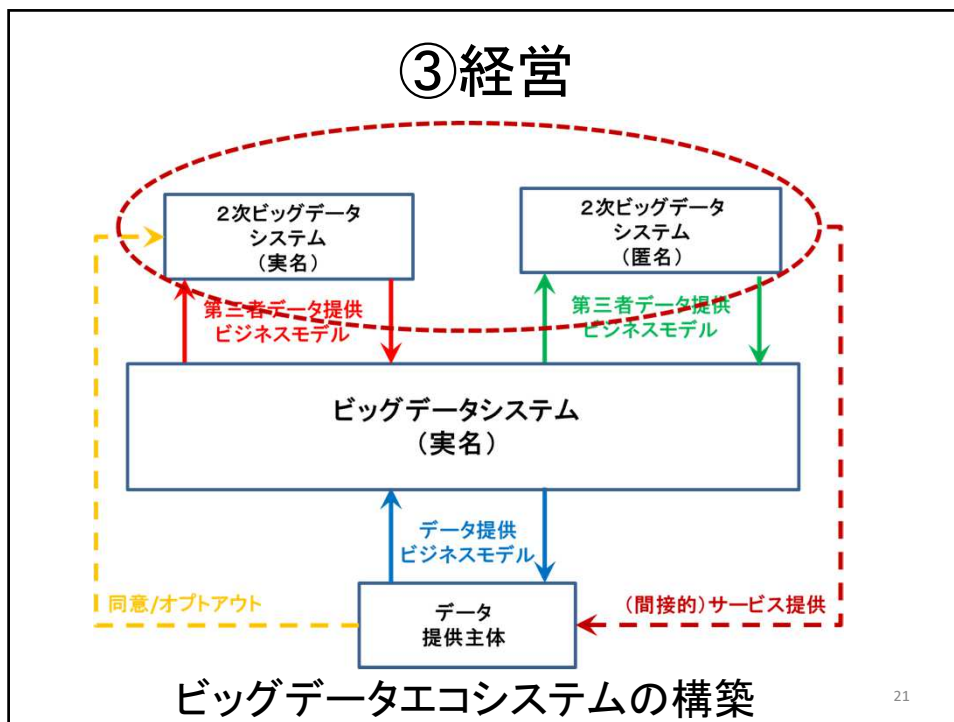
2014年の日本学術会議報告:

欧米やアジア諸国では統計学連する教育組織や学位授与数諸外国がデータサイエンティストの中で、我が国だけが逆に減少。今後の我が国の科学技術研究イノベーションにむけて重大な問題

米国	2万4730人
中国	1万7410
インド	1万3270
ロシア	1万2300
ブラジル	1万 90
ポーランド	8780
英国	8340
フランス	7770
ルーマニア	4970
イタリア	4900
日本	3400

データ分析専攻の学生、日本は育成に遅れ(データ分析の訓練を受けた大学卒業生の数)

科学技術白書(2016年文部科学省)²⁰より



④法制度

(1)ビッグデータの安心・安全のための法制度

(個人情報・パーソナルデータの適切な取扱い)

* 改正個人情報保護法(平成29年5月30日施行)

個人情報の定義の明確化(個人識別符号の定義)

公的な番号(マイナンバー、免許証番号等)

身体的特徴を示す符号

(顔、虹彩、指紋、DNA等)

慎重な取扱いを要する個人情報を定義

要配慮個人情報(病歴・犯罪歴に関するもの)

原則として本人の同意をとることを義務化

23

④法制度

(2)ビッグデータ活用促進を促す法制度

* 改正個人情報保護法(平成29年5月30日施行)

個人情報の事前の本人の同意を得なくとも

オプトアウト方式による第三者提供が可能

匿名加工情報の利活用に関する規定を新設

個人を識別できないように個人情報を加工し、

当該個人情報を復元できないようにした情報

第三者提供に

本人の同意、オプトアウトの仕組みは不要

* 次世代医療基盤法(平成29年4月成立)

要配慮個人情報に該当する医療情報も

オプトアウト方式により第三者提供可能

24

④法制度

(3)データの独占・寡占の弊害を防ぐ法制度

(自由な競争環境の確保)

データが大きな価値を持つようになりデータの独占や寡占が企業の競争を制限することにもなりかねないことが懸念

* OECD:「Big Data:Bringing Competition Policy to the Digital Era」(2016年10月)

* 公正取引委員会:「データと競争政策に関する検討会報告書」(2017年6月)

=>市場での支配的立場を使っでのデータ収集や
不当なデータの囲い込みは、独占禁止法の適用を検討

⑤合意(国民の理解)

(1)本人への直接的なサービスのためのデータ提供
データ提供主体へのインセンティブを確保し、
ビッグデータエコシステムを確保すること
=>個人を支えるビッグデータ

(2)社会を構成する国民の一人としてのデータ提供
国民生活を支える社会インフラの維持・高度化
国民の健康増進や医療・医薬の
改良・高度化のための基礎研究開発
=>社会を支えるビッグデータ(公共財?)

26

おわりに

- * ビッグデータの社会活用を推進する上で
想定される課題を整理した
- * 我が国が早期にビッグデータ社会を実現できるよう、
想定される課題を認識の上、
産官学それぞれの立場での活動が必要
- * 情報セキュリティ分野の技術者・研究者としても、
役割と責任を果たすべき課題も多い
- * 今後、暗号・認証技術の研究開発および実用化
を推進する立場から、技術・実用化課題を選定し、
その克服に向け注力予定

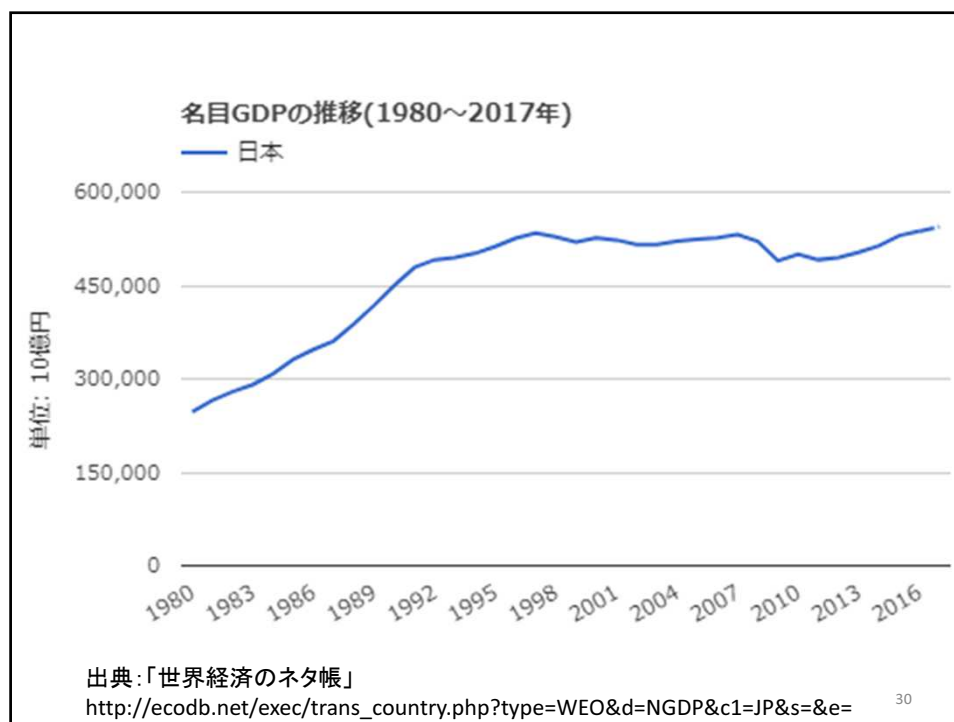
終

ビッグデータとは

「ビッグデータ(Big Data)」という用語の起源は2000年と言われているが、明確な定義は無い。

しかし、Gartnerのアナリストであるダグラス・レイニーはデータマイニングに関する2001年の報告書で、今後のデータ処理システムは情報資産(データ)を処理する量(Volume)と速度(Velocity)、および種類(Variety)という三つの軸(3V)での増大に対応できる必要があると主張している。

この3Vが一般にビッグデータの特徴とされている。²⁹



Society5.0とは

政府の総合科学技術・イノベーション会議が、2016年度から5年間の科学技術政策の基本指針となる「第5期科学技術基本計画」(2016年1月閣議決定)の中で、今後の重点テーマにしているコンセプト。

具体的には、狩猟社会、農耕社会、工業社会、情報社会に続く、以下のような新たな経済社会。

- ① サイバー空間とフィジカル空間を高度に融合させ、
- ② 地域、年齢、性別、言語等による格差なく、多様なニーズ、潜在的なニーズにきめ細かに対応したモノやサービスを提供することで経済的発展と社会的課題の解決を両立し、
- ③ 人々が快適で活力に満ちた質の高い生活を送ることができる、人間中心の社会(超スマート社会)。

31

第4次産業革命とは

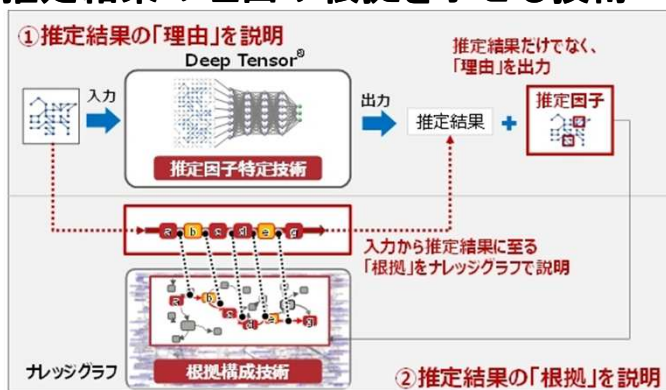
《Industry 4.0》ドイツの産官共同プロジェクトが提唱した、新たな産業高度化の概念。(2011年、SAP会長がドイツ工学アカデミー会長就任時に提唱、ドイツの国家PJに採用され、ドイツ国内で議論がスタートした概念。)

蒸気機関を第一次、電気機関を第二次、製造業の自動化を第三次の産業革命とみなし、インターネットを通じてあらゆる機器が結びつく段階(IoT)を第四次の産業革命と位置づける。

主に製造業を中心に、自律的・自動的・効率的に製造工程や品質の管理を進め、省エネルギー化などを行い、産業の高度化を目指すというもの。

32

AIの推定理由や根拠を説明する技術 <富士通が2017年9月20日にPress Release> グラフ構造のデータを学習する富士通研独自のAI技術「Deep Tensor」と、学術文献などの専門的な知識を蓄積したナレッジグラフの知識ベースを関連付けることで、AIの推定結果の理由や根拠を示せる技術

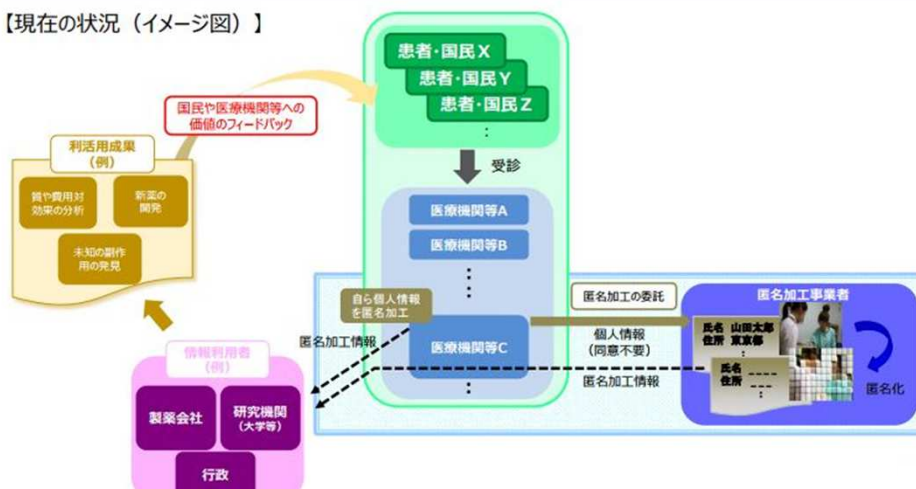


(Zdnet Japan 9月20日の記事より)

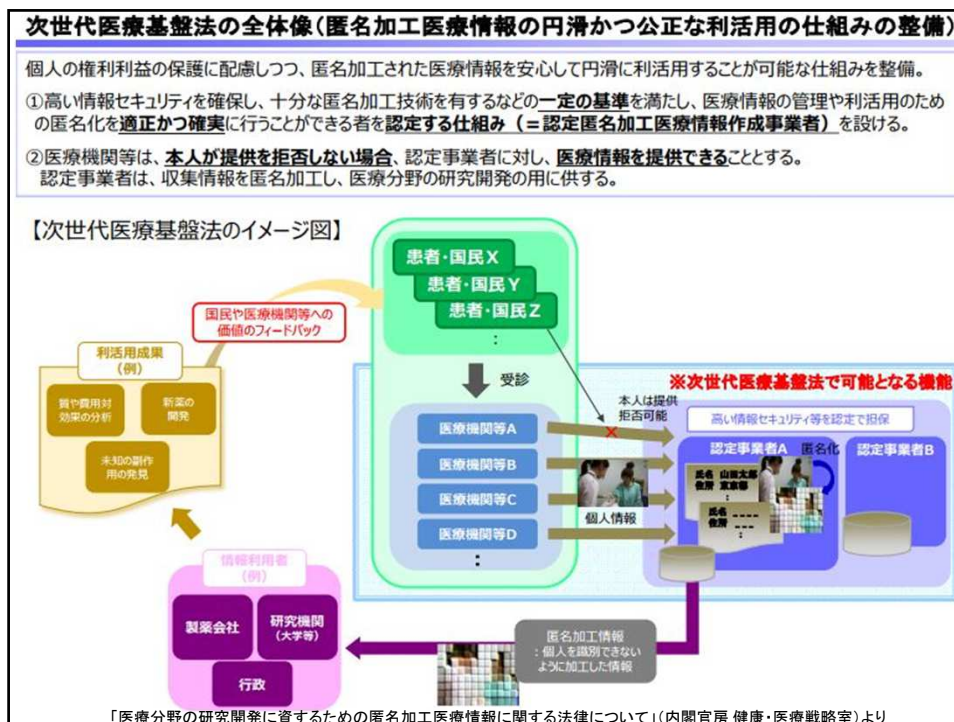
現行法で可能な匿名加工医療情報の提供の仕組み

- 匿名加工情報は、特定の個人を識別することができないよう個人情報を加工して得られる個人に関する情報であって、当該個人情報を復元することができないようにしたもの。
- 匿名加工情報については、本人の同意なく第三者に対する提供が可能。
- このため、個別医療機関は、保有する医療情報（個人情報）の匿名加工を自ら又は事業者等に委託して行い、利活用者に本人の同意なく提供することは可能である。

【現在の状況（イメージ図）】



「医療分野の研究開発に資するための匿名加工医療情報に関する法律について」(内閣官房 健康・医療戦略室)より



認定匿名加工医療情報作成事業者

認定候補組織：

- ①特定非営利活動法人日本医療ネットワーク協会
千年カルテPJの推進
- ②一般財団法人 医療情報システム開発センター
保健医療情報のシステム化推進
- ③日本医師会