

ビッグデータの 利活用

ビッグデータの有用性と課題

日本大学法学部 2013年度フォーラム発表

山田正雄ゼミナール 13期生

川瀬遼太郎 櫻井健 中森裕子 三好広和 元井佑也 山本浩大 吉田駿介

はじめに

はじめに①

- 近年「ビッグデータ」の利活用がさまざまな組織で検討、実施されている。平成25年度版情報通信白書では、ビッグデータの利活用による経済効果は年間約7兆7700億円に達するとの試算が出ている
- その利活用分野はマーケティングだけでなく、防犯や交通インフラの改善などにも及んでいる。ビッグデータによって、従来より質の高いサービスの提供ができ得ることから注目を浴びているのである

はじめに②

- 本研究ではまず、ビッグデータとは何かを明らかにし、誰が、どのように、何のために、利活用するのかを検討する
- そして、収集や分析といった利活用プロセスそれぞれで発生する問題点について、技術的、法律的、道義的な面から検討する
- また、個人データの利活用の際にはプライバシー侵害の恐れがある。特に第三者への個人データの提供が行われる場合の問題点について検討し、今後期待されるビッグデータの利活用とはどのようなものかを示していく

目次①

- はじめに
- 1. ビッグデータとは何か
 - 1-1. ビッグデータの3V
 - 1-2. ビッグデータを構成するデータの種類
 - 1-3. データ化する対象
 - 1-4. ビッグデータの登場背景
 - 1-5. まとめ
- 2. ビッグデータの利活用
 - 2-1. ビッグデータの利活用プロセス
 - 2-2. ビッグデータ利活用事例
- 3. 誰が、どのように、何のために
 - 3-1. 誰が
 - 3-2. どのように
 - 3-3. 何のために

目次②

- 4. 各プロセスにおける問題点
 - 4-1. 収集・蓄積する際に発生する問題点
 - 4-1-1. 技術的な問題点
 - 4-1-2. 法律的な問題点
 - 4-1-3. 道義的な問題点
 - 4-2. 分析・活用する際に発生する問題点
 - 4-2-1. 技術的な問題点
 - 4-2-2. 法律的な問題点
 - 4-2-3. 道義的な問題点
- 5. 第三者への個人データの提供
 - 5-1. 「個人情報」とは
 - 5-2. 個人情報の第三者提供とは
 - 5-3. 第三者提供になる場合
 - 5-4. 第三者提供にならない場合
 - 5-5. グレーゾーン
 - 5-6. 個人データの取り扱いに対する注意点
- おわりに

1.ビッグデータとは何か

1-1. ビッグデータの3V

- 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」
- 管理が困難になる要因は、**3V(Volume量/Variety多様性/Velocity発生頻度・更新頻度)**である
- つまり、従来よりも大量かつ種類が豊富なデータで、さらにデータが発生・更新されるタイミングが様々なため、従来のシステムでは蓄積・処理・分析が困難であるものを、ビッグデータと呼ぶ

3つのV	説明
Volume【量】	従来よりも、 <u>大量のデータ</u> を扱えるようになった。
Variety【多様性】	従来よりも、 <u>多様なデータ</u> を扱えるようになった。
Velocity【頻度】	従来よりも、 <u>頻繁に発生・更新するデータ</u> が扱えるようになった。

1-2.ビッグデータを構成するデータの種類

ソーシャルメディアデータ

例 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

例 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

例 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

例 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

例 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

例 オフィスのパソコン等において作成等されるオフィス文書、Eメール等



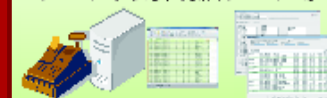
ログデータ

例 ウェブサーバー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

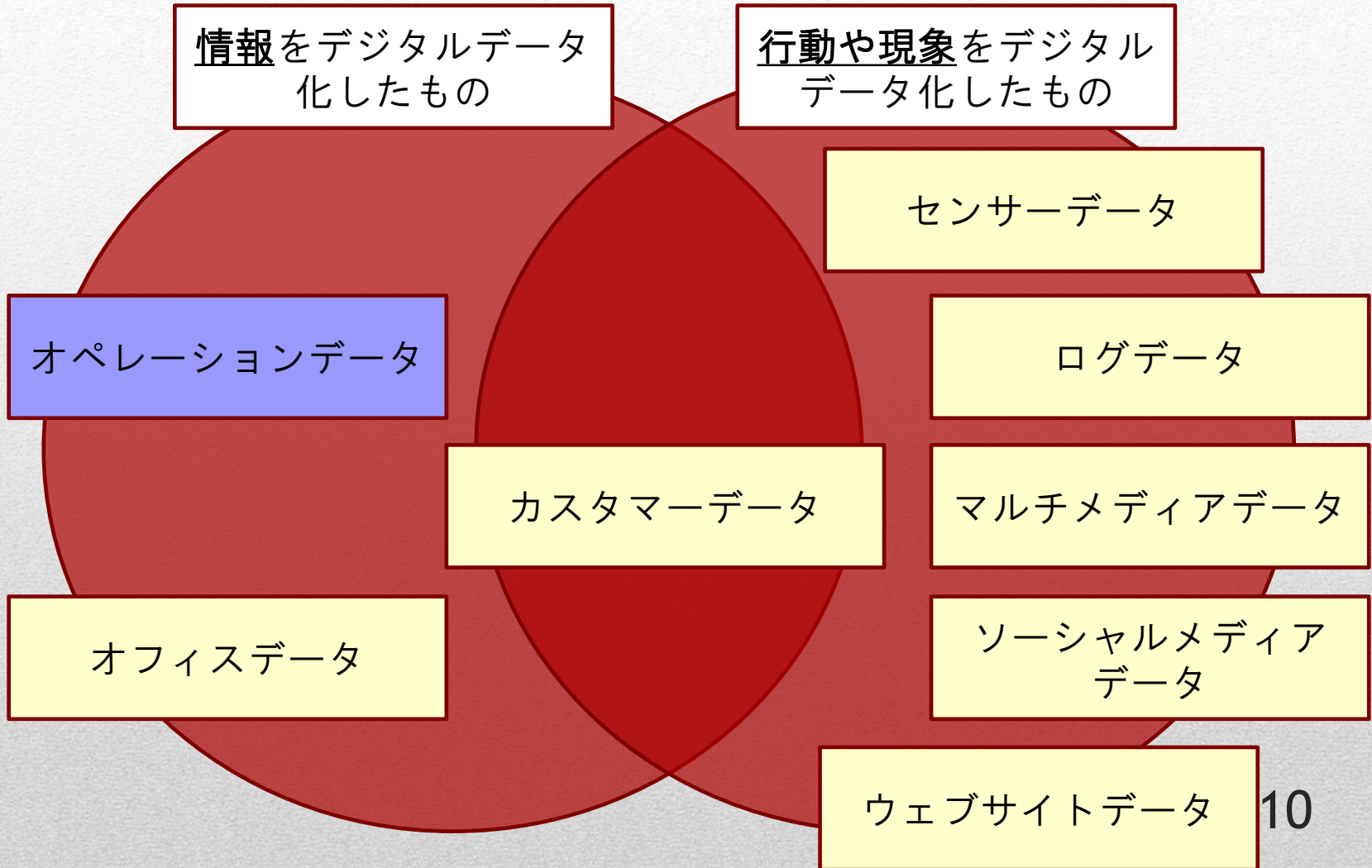


オペレーションデータ

例 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



1-3. データ化する対象



1-4.ビッグデータの登場背景

- 導入コストの低下

- ①データを蓄積するストレージの容量当たりの価格が、年々下落している
- ②コンピューターの処理能力が向上した結果、過去と比べて安価で高性能な設備が導入可能になった
- ③クラウドサービスが普及したことで、設備を自前で用意する必要が軽減された

- 3Vデータに対応した技術の登場

従来の技術では分析することが出来なかった、3Vデータの分析に適したHadoop等の技術が登場した

1-5.まとめ

- ビッグデータは、従来のデータ活用の延長線上にあるが、3Vという点でより有用性が高まっている
- ICTの発展により導入コストが低下し、分析技術が進歩したことで、これまで利用できなかったデータも利活用できるようになった
- これまで利用できなかったデータによって人々の行動や現象を分析できるようになり、それらを対象にした利活用が可能になった

2.ビッグデータの利活用

2-1.ビッグデータの利活用プロセス

- ①データを**収集**し、蓄積、結合する
- ②目的に合わせて抽出し**分析**(再結合)し、活用する



- ①、②のプロセスを踏まえることで、
★**サービスの向上**が可能になる

2-2.ビッグデータ利活用事例①

- Amazonによる、利益促進

①会員データ（顧客属性）、ウェブサイト上での利用者の閲覧履歴（クリックストリーム）、購買履歴等を**収集**

②同じ商品購入者の動向などを**分析**し、個人の購買行動までの流れの把握（何をみて、どんな行動をしてから購入行動に至ったのか等）

☆顧客の嗜好を把握し商品をおすすめし、**サイト上での購買行動を促進**

- Facebookによる、サービス利用促進

①会員データ、クリックストリーム、位置情報、利用者の書き込み等を**収集**

②会員の傾向を**分析**し、利用を継続する会員、退会する会員の行動を把握

☆会員に関連性の高い会員を「知り合いかも？」と表示するなど、**利用を促すサービスを提供**

2-2.ビッグデータ活用事例②

- コマツによる、建機の予防保守

①センサーで**収集**した建機の稼働状況データを、無線ネットワークを利用して自社サーバーに蓄積

②**分析**により故障の前兆現象のパターンを発見。故障のタイミングを正確に予測、予防保守が可能に



☆故障による業務の停止を防止、メンテナンス体制の効率化を実現

2-2.ビッグデータ利活用事例③

- 埼玉県と本田技研工業による、交通インフラ整備

①本田技研工業のカーナビゲーションシステム「インターナビ」で、道路走行データを**収集**

②埼玉県がデータを**分析**し、急ブレーキが多発する道路を**特定**

路面標示による注意喚起などの安全対策を実施



☆**急ブレーキ回数が約7割、事故件数が約2割減少**

2-3.ビッグデータ利活用事例④

- サンタクルーズ市による、犯罪の予防

①窃盗犯罪防止システムで市内のデータを**収集**

②窃盗事件についての過去8年分及び収集したデータの**分析**により、犯罪者の行動パターンや、店舗の営業時間等の環境要因と犯罪発生との関係性を発見



☆最も犯罪が起こる可能性の高い地域を予測し、犯罪が発生する前に現場に警察官が現れ犯罪を防止

システムを導入した2011年7月以降、**犯罪件数が減少**

3.誰が、どのように、何のために

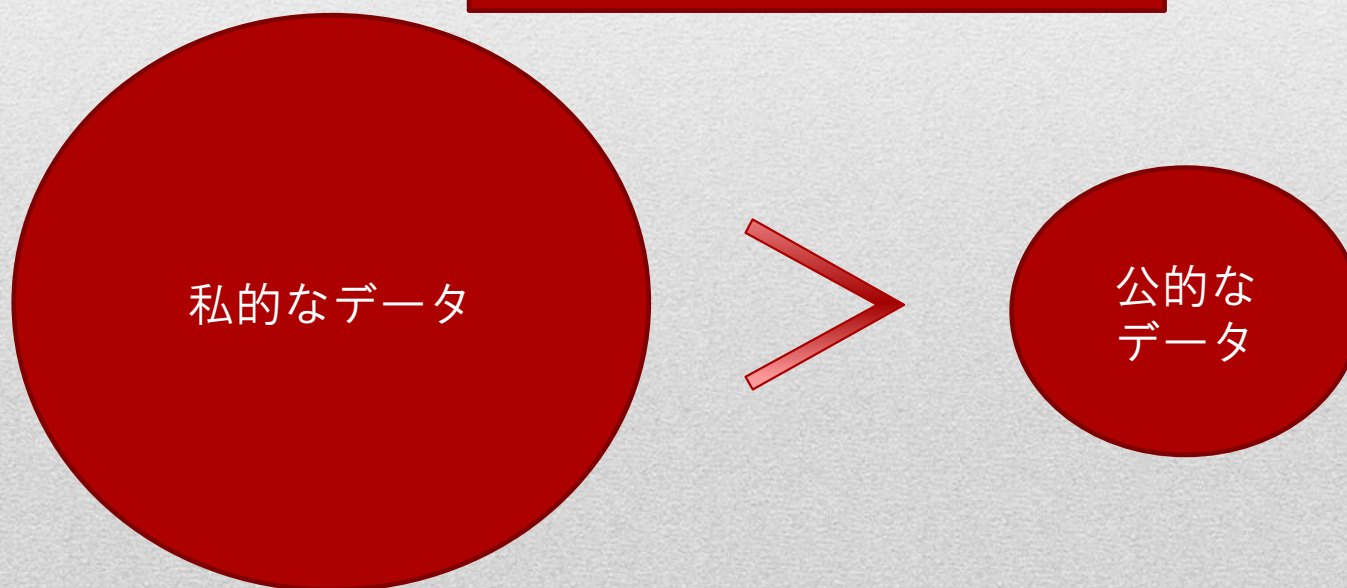
3-1. 誰が①

- 利活用する主体
 - ①公的な組織
 - ②私的な組織
- 公的な組織が持つデータ
 - 戸籍や年金記録など、公務に関するデータ
- 私的な組織が持つデータ
 - それぞれの組織が提供しているサービスの利用者に関するデータ。利用者の様々な経済活動や行動の履歴データ。購買記録、会員情報等

3-1. 誰が②

- 公的な組織が持つデータより、私的な組織が持つデータのほうが個人の性格・趣味・趣向を対象にしている(=個人に関する詳細なデータ)

個人に関するデータの詳細さ



3-1.誰が③

- 収集や分析等の各利活用プロセスを...

①自らで**全てこなせる**主体→FacebookやAmazon等

②自らで**全てこなせない**主体→その他多くの主体



- ほとんどの場合、一つの主体が利活用を完結して行うことができるわけではない

3-2.どのように①

①データの**収集**方法は、主体や利活用分野で異なる

- 自ら収集

- ・ FacebookやAmazonでは、**サイト上のユーザーの行動履歴(クリックした回数・場所、そのページを開いていた時間など)**を記録して収集
- ・ コマツでは、**建設機械にGPSやセンサー等**を取り付けて、機械の**現在位置、稼働時間、稼働状況、燃料の残量**などのデータを収集し、サーバーに蓄積
- ・ サンタクルーズ市では、**犯罪記録のほか、市内の店舗や施設の運営情報を窃盗犯罪防止システム**で収集

- 他の主体と提携して収集

- ・ 埼玉県の事例では、**本田技研の車体に取り付けたセンサー**によって、データを収集

※そのほかに、自社で収集したデータと他社で収集したデータを組み合わせたり、他社に収集を外注する場合も考えられる

3-2.どのように②

②データの**分析**方法も、主体や利活用分野で異なる

- 自ら分析

Facebook、Amazon、コマツ、サンタクルーズ市の事例では、**自ら開発したシステム**で分析

- 他の主体と提携して分析

ホンダ技研と埼玉県は、**本田技研とIBMが提携したシステム**で分析

※そのほかに、自社で分析したデータと他社で分析したデータを組み合わせたり、他社に分析を外注する場合も考えられる

3-3.何のために

☆サービスの向上という目的も、主体や利活用分野で様々である。

- FacebookやAmazon、そしてコマツでは、マーケティング促進手段の一つとして活用している
- サンタクルーズ市では、治安維持や犯罪予防に活用
- 本田技研と埼玉県は、埼玉県の交通インフラの改善のために活用

4.各プロセスにおける問題点

4-1.収集・蓄積する際に 発生する問題点

4-1-1.技術的な問題点

- サードパーティのクラウドサービスを蓄積環境として利用する際の問題点
- プロバイダーが破産や操業停止するとデータにアクセスできなくなるリスクがある



- 機密情報や利用者の詳細な個人情報等の重要なデータは自社でバックアップを取り、いざという時のデータ消失に備える必要がある。

4-1-2. 法律的な問題点

- プライバシーを侵害する違法な収集

イギリスでは、Googleが「Street View」撮影車両によってセキュリティ対策の不十分なWi-Fiネットワークから個人情報データを収集した行為について、同国のデータ保護法に違反していると認定



- コンプライアンスを徹底し、社会的責任(CSR)を全うしなければならない

4-1-3.道義的な問題点①

- 共通ポイント「Tポイント」を運営するカルチュア・コンビニエンス・クラブ（CCC）
 - ・ 加盟している企業間で、利用者情報を共有してサービスを提供している
 - ・ 加盟店で集められた購買情報等は、CCCに一元管理されている



加盟店で集められた情報と、氏名や住所などの個人情報
が統合されると、誰が何を買ったのか、こういった嗜好
を持つのか等が特定可能になる

4-1-3.道義的な問題点②

- 個人情報保護法

第二十三条 「個人情報取扱事業者は、(中略)あらかじめ本人の同意を得ないで、個人データを第三者に提供してはならない」

- 経済産業省「個人情報の保護に関する法律についての経済産業分野を対象とするガイドライン」

共同利用が合法となる場合

- ①共同して利用される個人データの項目
- ②共同利用者の範囲
- ③利用する者の取得時の利用目的
- ④安全管理等個人データの管理について責任を有する者の氏名又は名称

これら4つを、「あらかじめ」利用者本人に通知しなければならない

4-1-3.道義的な問題点③

- 保護法でもガイドラインでも、前もって利用者本人の同意を取れば、収集や共有は可能としている
- 利用規約の中で「個人情報項目を共同して利用」とすると明記

法やガイドライン
に
反してはいないが

- 利用規約の文字数(約一万字)や読みづらさを考慮すると、明記しても読まれない(理解されない)可能性がある

4-1-3.道義的な問題点④

- 利用者が規約内容を理解したうえで同意するためには？

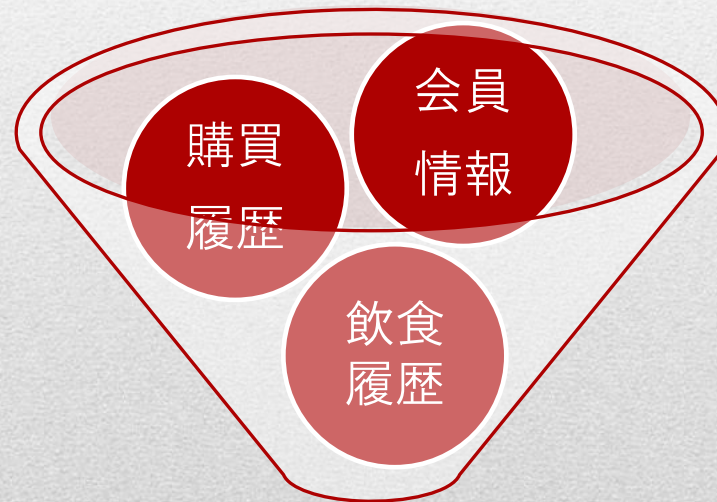
解決策

- ①規約の内容について別途簡潔な説明を用意
 - ②イラスト等、分かりやすい利用ガイドの作成
- 事業者は重要事項について利用者へ通知し認識させる必要がある

4-2.分析・活用する際に 発生する問題点

4-2-1.技術的な問題点①

- 個々のデータは詳細なものだけでなく、他のデータとの統合により、具体的に個人の趣味や嗜好を把握できるデータが生成される



詳細な個人情報

4-2-1.技術的な問題点②

- ビッグデータは個人の詳細なデータ
- 分析や加工等、データの利用において、従来よりも取り扱い段階でのリスクが伴う



解決策

- 情報を保護し漏えいを防ぐ等、よりプライバシーに配慮したシステム構築が必要

4-2-2. 法律的な問題点

- 最もプライバシーを侵害するプロセスは、活用され様々なサービスとして利用される段階である。利用者に影響を及ぼさなければ、実害は出ない(もしくは少ない)



プライバシー侵害を防ぐためには？



- 違法なデータの利活用を防ぐためには、その前段階である分析段階でプライバシーを侵害していないか再確認する必要がある

4-2-3.道義的な問題点

- 業務、人材管理での活用

自社の従業員の働きぶり(勤務時間や勤務中の行動等)を分析し、事業の効率化を図る

こうした分析が人事評価に活用された場合、**従業員の行動を束縛する、監視システムになりうる可能性あり**



解決策

- 業務にデータ分析を用いる際には、従業員のプライバシーを考慮する必要がある
- 人事評価に利用する際には、その根拠となった基準やデータを、評価の対象となった本人に明示し同意を得る必要がある。

5. 第三者への 個人データの提供

- 必ずしも一つの主体がビッグデータの利活用プロセスすべてを実行するわけではない



- 多くは第三者がこのプロセスに介入している



- 第三者が介入することによってどのような問題点があるか？

5-1.「個人情報」とは

- 個人情報保護法 第2条

この法律において「個人情報」とは、生存する個人に関する情報であつて、当該情報に含まれる氏名、生年月日その他の記述等により特定の個人を識別することができるもの（他の情報と容易に照合することができ、それにより特定の個人を識別することができることとなるものを含む。）をいう。

5-2. 個人情報の第三者提供

- 第三者提供をする場合

あらかじめ活用目的を明確に示してから…



- ① それに対して許諾を取ってから提供する
→ オプトイン方式
- ② 許諾なしで提供できるが、利用者からの要請があれば当該データを削除し利活用しない
→ オプトアウト方式

5-3. 第三者提供になる場合

- グループ会社間で個人データを活用する場合
- 他の組織間（フランチャイズ組織の本部と加盟店、同業者間等）で個人データを活用する場合



- あらかじめ活用目的を明確に示し、オプトイン方式かオプトアウト方式を設ける必要がある

5-4. 第三者提供とはならない場合

- 社内で利活用する場合



- 同一組織内での情報利用は、第三者提供にはならない
- あらかじめ活用目的を明確に示して同意をとる必要はない

5-5.グレーゾーン①

- 共通ポイント「Tポイント」を運営するカルチュア・コンビニエンス・クラブ（CCC）
 - ・ 加盟している企業間で、利用者情報を共有してサービスを提供している
 - ・ 加盟店で集められた購買情報等は、CCCに一元管理されている



- 加盟店で集められた情報と、氏名や住所などの個人情報
が統合されると、誰が何を買ったのか、こういった嗜好
を持つのか等が特定可能になる

5-5. グレーゾーン②

- 加盟企業間で、共通のポイントカードを使用している場合
- ポイントカード情報を加盟企業間で利用してもよいのか



- こうしたポイントカードシステムを、社内としてみなすのか、それとも第三者提供とみなすのか、法令やガイドラインが追いついていない現状

5-6.個人データの取り扱いに対する注意点①

①自ら収集する場合

個人からデータを収集する際には、データ本来の持ち主に対して、利用の目的を必ず事前に個別的かつ明示的に説明しなければならない

そのデータを収集した経緯や目的に沿った使い方をしない場合は、改めて同意を得る必要がある

②他の主体から提供を受ける場合

データを購入等して収集・分析するケースでは、そのデータの取得方法や利用許諾の有無を確認する必要がある

5-6. 個人データの取り扱いに対する注意点②

③ 自社でデータを分析/第三者がデータを分析する場合

収集したデータに対するアクセス権限や取り扱いに注意し、第三者に加工を依頼する際には、プライバシーマークなど提供先の体制を確認し情報漏えいリスクを減らす必要がある

④ データを提供する場合

膨大なデータは汎用的な価値を持つため、データを持つ主体は提供や販売を請われるケースが今後出てくると考えられる法令遵守と並行して、利用者への事前説明と、何が個人情報にあたるかを検討したうえで販売を行う必要がある

おわりに

おわりに①

- ビッグデータの利活用において、他の主体の様々なデータを活用できることには非常に有用性がある。今後、ビッグデータはその有用性ゆえに、企業はもちろん、政府・官公庁においても利活用がより推進されていく
- しかし、利活用のプロセスやデータの第三者提供といった面において問題のある事例も見受けられた。ビッグデータの利活用に関わる各主体には、どのような展望があるだろうか

おわりに②

- 個人

ビッグデータによる便利なサービスを享受できる主体であるのと同時に、データの収集対象であるがゆえにプライバシーの侵害を被る主体でもある。利用規約内容を事前に把握しておくなど、自らのプライバシーを守ることも必要だが、データの収集が多様になればなるほど徹底は難しくなっていく

- 企業

利活用の有用性を促進するため、あらゆる方法でできる限り個人情報収集・活用していただく。便利な利活用と問題のある利活用が、今後ますます増加すると考えられる。企業がCSRの全うやコンプライアンスの遵守をすることは必要だが、それだけでは不十分

- 政府

個人のプライバシーを守るため、増大した事例をふまえて法令やガイドライン等をより現状に即したきめの細かいものとし、問題のある利活用を監督する役割が求められる

参考文献

- 『一時間でわかる 図解ビッグデータ 早わかり』 大河原 克行 中経出版 2013年
- 『情報通信アウトLOOK2013ー ビッグデータが社会を変える』 情報通信総合研究所 NTT出版 2012年
- 『データサイエンティスト完全ガイド』 日経情報ストラテジー 日経BPムック 2013年
- 『ビッグデータの衝撃』 城田真琴 東洋経済新報社 2012年
- 『ビッグデータの正体』 ビクター・マイヤー=ショーンベルガー/ケネス・クキエ/斎藤栄一郎訳 講談社 2013年

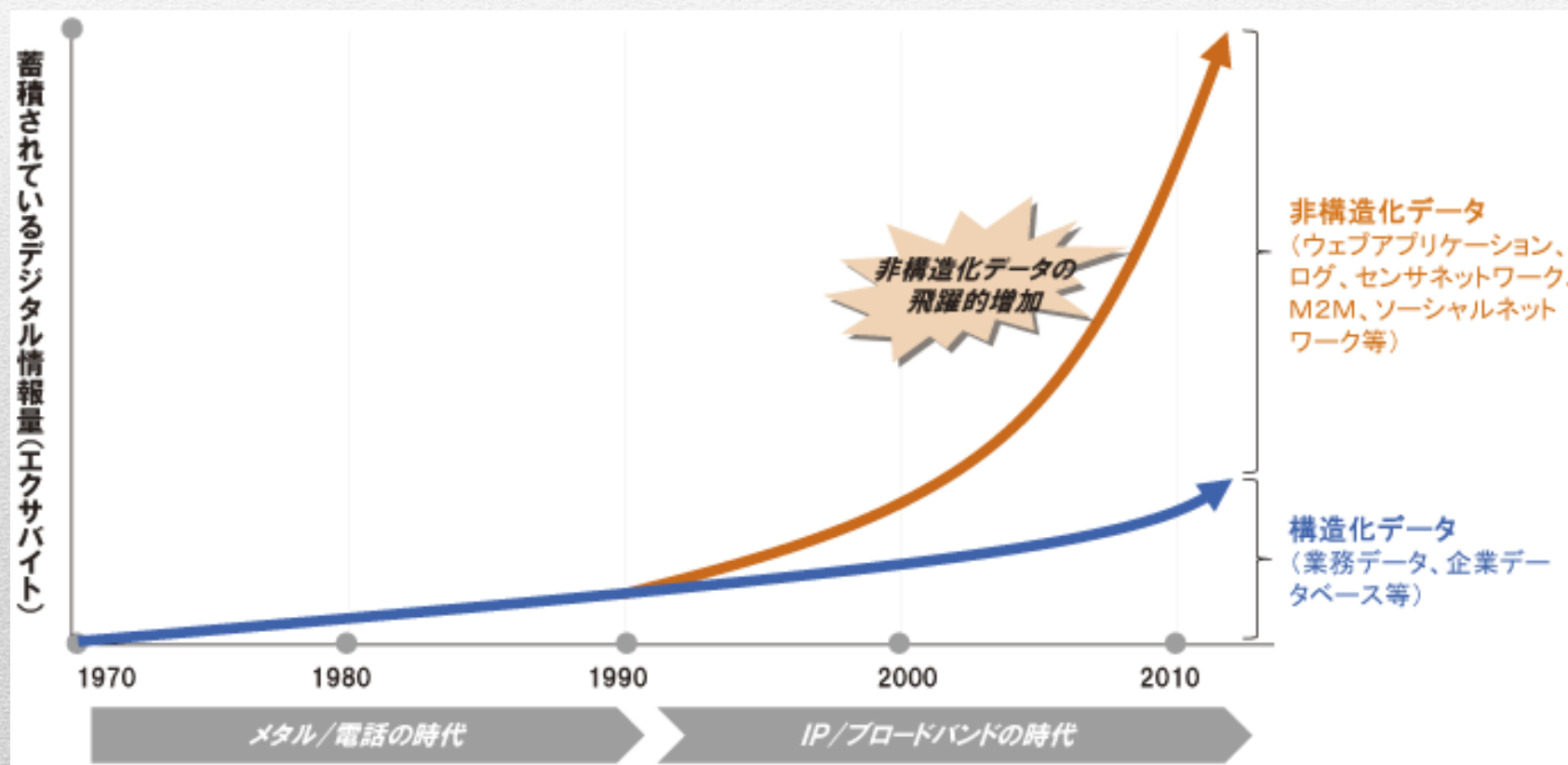
参考URL

- 旭化成株式会社 <http://www.asahi-kasei.co.jp/asahi/jp/index.html>
- EMC ジャパン <http://japan.emc.com/index.htm>
- 総務省 <http://www.soumu.go.jp/>
- NHK ニュースweb <http://www3.nhk.or.jp/news/>
- NTT ドコモ <http://www.nttdocomo.co.jp/>
- KDDI 株式会社 <http://www.kddi.com/>
- JR東日本 <http://www.jreast.co.jp/>
- 日本IBM <http://www.ibm.com/jp/ja/>
- 日本経済新聞 <http://www.nikkei.com/>
- 日本電気 NEC <http://jpn.nec.com/>
- 日立製作所 <http://www.hitachi.co.jp/>
- ビッグデータ・オープンデータ活用推進協議会
<https://www.facebook.com/bigdataopendata4city>
- 富士通 <http://jp.fujitsu.com/>
- 法令データ提供システム <http://law.e-gov.go.jp/htmldata/H15/H15HO057.html>
- 本田技研工業株式会社 <http://www.honda.co.jp/>

補足資料

補足①:ビッグデータのVolumeの増加

- 構造化データだけでなく、非構造化データも扱うことで、従来のデータ分析よりも大量で多様なデータを扱うようになってきている



補足②：構造化データとは

構造化データとは、ある決まった枠組み(フィールド)の中に納まるデータのこと

具体例：エクセルのセルに入力し、計算やグラフ化できる**数値**など。

これまでデータ分析と呼ばれていたものは、こうした構造化データのみを対象としていた

主な構造化データ	説明	利活用の目的
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。	こうしたデータの分析の目的は、発注や在庫管理の効率化、売れ筋商品や販売傾向の発見など。

補足③：非構造化データとは

非構造化データとは、そのままでは構造化データのように整理・分析できない、多種性とリアルタイム性があるデータ
文章・音声・画像・動画・SNSの投稿・各種センサーデータなど多岐に渡る

非構造化データの特徴	説明
多種性	構造化データに比べ、種類が多い。割合は、2:8程度と言われている。
リアルタイム性	データの生成・収集・蓄積・分析、そしてその結果を利活用するタイミングが、リアルタイムであること。

補足④.利活用分野

- 金融 自動車走行実績に基づく保険商品の販売
- 農業 栽培条件データを用いた作物品質管理、衛星写真を用いた作物生育状況監視、植物工場
- 製造：遠隔監視データを用いた新たなサービス展開
- 流通：ソーシャルメディアを用いた販売促進、位置情報を用いた販売促進
- 設備管理：ビルや住宅のエネルギーマネジメントシステム電力の使用状況を蓄積、分析することで利用状況に応じた最適な空調、電気製品の制御等を可能とし、電力消費量の削減等に寄与
- 観光：KDDIと三重県伊勢市、埼玉県、徳島県が、位置情報ビッグデータを地域振興に活用