資料基７－２
（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

目次

第１回会合（平成24年2月7日）関係

１．本田技研工業（株） ・・・Ｐ　３

２．富士通（株） ・・・Ｐ　７

３．ＫＤＤＩ（株） ・・・Ｐ１５

第２回会合（平成24年2月20日）関係

４．東京海上日動火災保険（株） ・・・Ｐ２０

５．（株）ウェザーニューズ ・・・Ｐ２４

６．日本アイ・ビー・エム（株） ・・・Ｐ２９

第３回会合（平成24年3月8日）関係

７．（株）プリファードインフラストラクチャー ・・・Ｐ３３

８．日本電信電話（株）／
（株）エヌ・ティ・ティ・ドコモ ・・・Ｐ３７

第４回会合（平成24年3月23日）関係

９．（株）エヌ・ティ・ティ・データ ・・・Ｐ４５

第５回会合（平成24年4月13日）関係

１０．インテル（株） ・・・Ｐ５２

その他国内外における取組事例

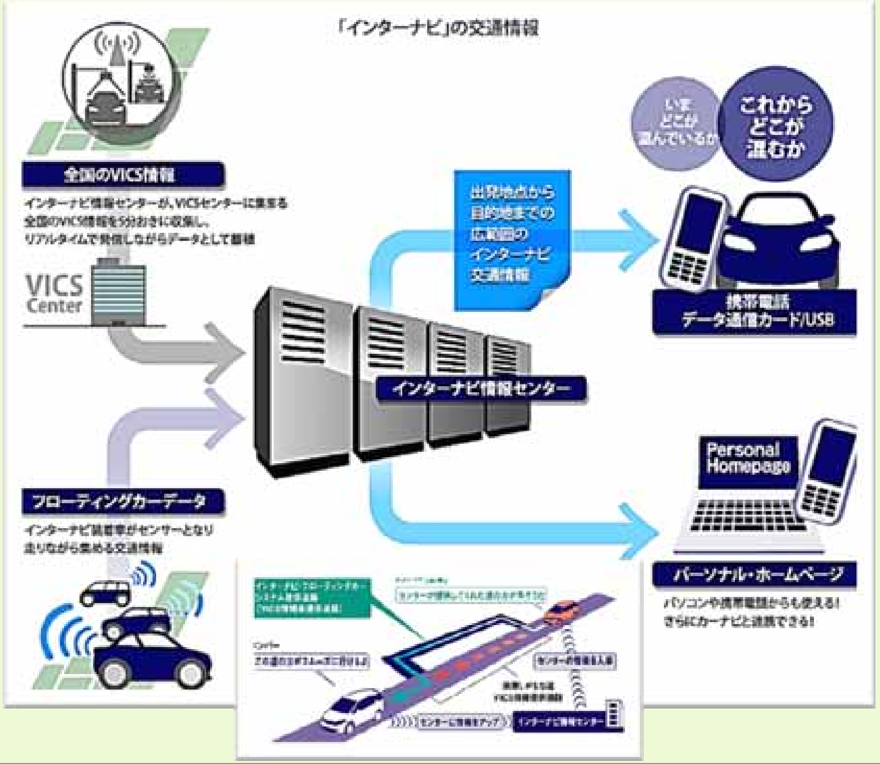
１１．国内におけるその他取組事例 ・・・Ｐ５５

１２．海外における取組事例 ・・・Ｐ５７

1．本田技研工業（取組事例）

取組事例の概要（1／2）

- ドライバーの快適なカーライフを実現するため、より安全で環境にも配慮したドライブ情報サービス・ネットワークとして、安全・安心、防災、天気、省燃費ルート等の情報を提供する「internavi」が2002年からサービス提供開始。現在、会員数は145万人。
- 2003年に、会員から５分毎等の間隔で収集した「internavi」装着車の走行データ（Floating Car Data：FCD）の共有により、渋滞を回避し、目的地へより早いルート案内を行う「フローティングカーシステム」を導入。主要幹線を対象とするVICS（Vehicle Information & Communication System）を補完し、現在、毎月約１億kmのデータがアップ中。2011年8月現在、蓄積した走行データは15億km（地球約37,500周分）。
- 以上のシステムによる効果としては、例えば、次のとおり。
  ①　VICSとFCDを融合した渋滞予測による効果検証結果（東名阪100サンプルのシュミレーション）では、約20%早いルートが案内され、CO2換算では約16%の削減効果。
  ②　埼玉県の道路行政の取組において、FCDより急ブレーキポイントを抽出し、街路樹の剪定や路面表示により、急ブレーキ回数が約７割減少。
  ③　2006年のＮＰＯ法人防衛推進機構の研究への協力により、2007年の新潟中沖越地震・2008年岩手宮城内陸地震において、FCDを活用した通行実績マップを生成・公開。また、昨年3月の東日本大震災においては、同マップをGoogleへ提供し、ＮＰＯ法人ITS-Japanをはじめ、各種行政機関や研究機関でも活用。



1. 本田技研工業（取組事例）

取組事例の概要（２／２）

- 東日本大震災での取組としては、以上の他、①津波警報と地震震度情報や首都圏の通行止め状況のカーナビ画面への配信、②地震時の位置情報付きの家族へのメールによる安否連絡、③国土交通省河川局が設置した11カ所の浸水センサー観測値のカーナビやスマートフォンへの配信等が実施。
- 目的地への走行ルート等と(財)日本気象協会から提供される凍結予測等の気象情報との連携により、路面凍結発生の予測時刻や予測地点等のカーナビ画面への表示・音声警告等の気象・減災情報を提供。
- 車両内のセンサーから収集している燃料噴射量のデータの活用により、例えば、燃費のよいルートの探索・予測や車両の制御等が可能。また、アメリカの車種ACURAでは、車載のCAN（Controller Area Network）による故障データ等の活用により、故障部位のディーラーや顧客への連絡等を実施。
- カーナビを装着した車には、通信モジュールが標準装備され、現在、９車種まで拡大。車検時（２年毎）にHonda販売店で更新手続をすれば、翌２年間の通信費がメーカ負担となり無料化。

1. 本田技研工業 (ヒアリング)

ヒアリングの概要（1／2）

☞ 車にはラジオやテレビが標準装備。この点、東日本大震災において、津波発生地域にいた「internavi」会員（109名）のうち74%が車内のラジオ等により大津波警報を認知。車内に標準装備のラジオ等に対し、情報を的確にいつでも送れることが大事であり、マルチメディア放送（V－LOW帯）について、地方におけるインフラ整備や普及が重要。

☞ 走行データの送受信については、携帯電話との有線接続やBluetoothによる無線接続等、携帯電話ネットワークを活用している。この点、対応可能な車種は増加しているが、通信手段として、ｉＰｈｏｎｅ等のスマートフォンには対応ができていないため、テザリング等が必要。

☞ 例えば震度６になる瞬間に、公共的な情報プラットフォームに各者からのデータが自動的にアップされ、必要なデータがサーバ間で連携・加工されて必要な情報の形になり、各ユーザや車に送信されるという仕組みが必要。その際、ＰＤＦ等ではない形式で、デジタルデータとしての情報プラットフォームが必要。

☞ 東日本大震災における通行実績マップの生成において、通行データを自動でＫＭＬファイルに変換可能な者と自動変換が不可能であり手作業で変換する者がいた。この点、前者のサーバに後者のデータが集まれば、自動変換が可能となるが、平時における後者による通行データの収集状況が明らかになり、前者に対して後者からデータが提供されなかったため、ＮＰＯ法人であるＩＴＳ－Ｊａｐａｎに対して、各者が通行データを提出した。しかしながら、手作業で変換して提出等する必要があったため、自動的な連携が可能となる体制等が必要。

☞ データの収集において、例えば、浸水センサーについては、管理者が様々であり、国土交通省河川局のほか、都道府県や市町村も管理しているため、それぞれとの調整が必要。車が走行する上で、土砂崩れの可能性等の重要な情報があるため、このような情報はできる限りオープンにし、プラットフォーム上に集めることが重要。

1. 本田技研工業（ヒアリング）

ヒアリングの概要（２／２）

☞　車の買い換えも減少し、ワンオーナーで９年になり、また、スクラップになるのが１３年という状況の中で、スマートフォンをそのままカーナビとして使用する等、カーナビをアップデート可能なものとする等の仕組みが必要。

☞　カーナビが進化する中で、安心・安全の観点において、情報の信頼性と裏腹の問題として、例えば、配信した「停まれ」という情報が誤っていた場合における責任追求の問題が、システム開発等を躊躇させるおそれが存在。

☞　１～２年で買い換えとなる携帯電話と異なり、カーナビは長期間搭載されため、セキュリティホールへの対応等とともに、製造物責任法（ＰＬ法）への対応も必要となる。このような中で、例えば、通行実績マップのように社会に貢献する情報をできるだけ提供等することの重要性とのバランスを踏まえ、免責事項（disclaimer）とした上で、最終的には運転手等の判断で情報が活用される形で提供。

☞　どういう方法により、どういう情報を提供等すれば避難や減災ができるのかについて、最後は運転手の判断になるが、その自助力を援助するためには、提供等される情報は的確にすることが必要。

☞　走行データ等の活用については、１００億円近い投資による１０年間の研究やサーバ等の構築等、民間企業においては、いろいろなデータを基に平時は競争している。例えば、ディーラーの出店場所について、どこが交通量が多く、道路のどの側に配置すると車が入りやすくなるのかの検討等に活用が可能となる。これらのデータを平時から公開等する場合におけるビジネスモデルや先行者利益の在り方が重要。この点、例えば、公共車両、トラック、タクシーや被災現場に向かう車等からスマートフォンでデータを収集することも必要。

☞　社会に役に立つ情報であり、そして、情報を提供する会員にもメリットがあることが大事。

2．富士通（取組事例）

取組事例の概要（1／2）

- 通信ネットワークを通じた人を中心とした新しい社会たる「ヒューマンセントリック・インテリジェントソサエティ」を目指し、同社会では、リアルワールドにおいて各種センサーからネットワークを通じて様々な情報が収集され、バーチャルワールドにおいてクラウド基盤等によりデータの融合や分析が行われ、知恵を組み合わせてリアルワールドにナビゲーションするサイクルが重要。このような「コンバージェンスサービス」により、企業や地球規模の課題を解決し、豊かな社会の実現を目指す。

- クラウド基盤の上に「データ活用基盤サービス（PaaS）」を本年から提供し、ここに蓄積される様々なデータを活用して、①顧客がここでアプリを作り、その最終ユーザに提供する「インテグレーション型」、②蓄積したデータを活用したアプリを作成して顧客に提供する「アプリ・サービス型」、③顧客が入力するデータを分析し、その結果を顧客に提供する「データ型」のサービスを実施。

- 「データ活用基盤サービス（ＰａａＳ）」の特徴としては、①複合イベント処理、並列分散処理、圧縮・秘匿化等の必要な技術を全て統合している点、②センサーデータやテキストデータ等の非構造化データ等の異種情報を多目的に扱う点、③クラウド上での提供により小さく始めて必要なだけ使える点の３点。

- データの分析等ができる専門家集団「キュレーター」により、データを持っ　ていない顧客やデータの活用方法等がわからない顧客等に対し、コンサル含め、データの有効性評価・分析やシステムデザインをする「データキュレーション」サービスを本年４月に体系化してサービス化予定。

2．富士通（取組事例）

取組事例の概要（２／２）

- ビッグデータの収集と利活用について、発生するデータの一部しか活用しておらず、例えば、次のような新たなサービスには活用しきれていないのが現状。
  ① 複合機や道路照明について、何か発生した際に収集するエラーデータの解析に加え、内部のセンサーからの膨大なデータをネットワークを通じて収集し、機器開発や故障前のメンテナンス等の新サービス。
  ② エアコンやテレビ等をネットワークでつなぎ、常に消費電力を管理し、外の気温と家中の気温により制御すること等の新サービス。

- 実験フェーズとしては、例えば、次の取組を実施中。
  ① コンビニ等に来店したが購入しなかった客の理由を知るため、ＰＯＳ情報による売上データ、店内カメラ画像解析による客の動線データ、そして、性別・年齢・気温等のデータを活用して、困っている可能性が高い等の顧客心理を分析し、客が購入しなかった場合の機会損失の低減や、店の棚割りや新商品の配置、レイアウトの改善等に関する企画・施策の適正評価を実施。
  ② 高齢化するベテラン農業従事者の知恵を共有するため、センシング技術等で圃場の野菜等の生育状況、気象情報や土壌情報を収集し、野菜等の市況情報、そして、ＧＰＳ携帯による現場作業者の作業・施肥実績ログ情報を活用して、工程管理、圃場管理、営農管理及び見回り支援により、食の安全・安定供給の確保、後継者の育成、品種改良の効率化等を実施。
  ③ 健康を促進しつつ医療費を削減するため、従業員2,500人規模を対象としたレセプトデータ、健康組合が持つ健康診断データ、体重や血圧等のバイタルデータを分析し、生活習慣病予備軍の早期発見や保健指導員による適切なアドバイスによる健康改善等の実証実験を実施。
  ④ 一般的な糖尿病の判定項目である空腹時血糖等の２項目を含め、中性脂肪等の全30項目について機械学習による総合判定を実施し、糖尿病になる可能性があり、実際に翌年になった人等について、機械学習による総合判定で高精度に予測が可能。
  ⑤ 活力ある街づくりのため、一部サービス提供中の「SPATIOWL」により、移動体情報の蓄積等による道路交通量等に基づく道路拡張等への反映、また、位置情報にプローブ情報等を重ねることによる渋滞情報等のプローブデータ活用サービスや通過可能エリア等の社会インフラ情報提供サービスを提供。

ヒアリングの概要（１／３）

☞　タクシープローブのデータは、タクシー事業者が配車業務で利用しているデータをそのまま使っているため、プローブのために自ら特別なネットワークを構築したり、特別な車載装置は使っていない。リアルタイムに変化するタクシーの位置を もとに、分析・加工し、新たなサービスとして顧客に提供。特に交通情報については、１日の走行距離が多いタクシーを利用して生成している。データの種類を増やすためには、乗用車のデータも融合できる可能性はあるが、現在は、実施していない。 また、どこかの自動車メーカと協力して標準化するという動きはしていない。

☞　クラウドサービスのベンダーとして、いろいろな分野でサービスを提供しており、異なる業種同士を結びつけて
新しいサービスをするための働きかけをしているが、成果はこれからである（実証実験中）。

☞　情報の秘匿化のため、暗号化と独自のフェデレーション技術を利用している。ビッグデータを活用するためには、個人を特定できない匿名化技術が必要。また、個人情報を、個人が特定できないデータとして、統計や計算などに活用する場合、可逆不可である匿名化が必要。また、匿名化であっても、複数のデータで個人が特定化できるものは、個人データの扱いとなるため、個人データを匿名化して扱うには、個人が特定できないＫ－匿名化の実施が必要。しかし、Ｋ－匿名化は、すべての組み合わせをチェックする必要があるため、計算に時間がかかり、ビックデータを扱うには課題。また、これにより、健康保険向けのデータ分析サービス等を実施するためには、業務委託契約が必要であり、広く蓄積データを活用できないことも課題。

☞　店内カメラの映像のマーケティング目的の解析については、現在は実験段階で実用化はしていないが、運用上の工夫で個人を特定しにくいような映像とするように検討中。

2．富士通（ヒアリング）

ヒアリングの概要（２／３）

☞　農業従事者や医師の知恵をシステムに落としこむ際の阻害要因として考えられるのは、以下のとおり。

＜農業＞

①　データ収集のための生産者が入れやすい簡易なユーザーインターフェースが必要。現在はＧＰＳやセンサーを使っているが、音声・映像等の検索やディクテーション能力がソフトウェア側に不足しており、蓄積したデータを活用する方法がまだ乏しい。

②　ユーザ（特に生産者）へのデータ活用の価値説明がしきれていない。

③　地域に散在する農業技術・知識を基盤として整備することが必要。１社単独の投資では面的展開は困難であり、地方自治体や国による支援を期待。

＜在宅医療＞

①　医療、介護、看護、薬局などでのルールの違いや、情報リテラシーの違いなどが存在するため、多職種間での情報共有においては入力や表示等ついて工夫が必要。

②　院長が高齢化しているケース等では、電子カルテや在宅医療クラウドなどのＩＣＴ化への理解を得ることが困難なのが現状。

③　個人情報の取扱いや、医療データの共有化の壁、多職種間での標準指標のオーソライズ、医療系クラウドに対する規制緩和等が課題と考えられる。

☞　データの有効性評価や収集・活用に向けた提言、指標化等を行うキュレーターは、現在は２ケタの人数規模で、今後３年以内に３ケタにする予定。主に以下のバックグラウンドをもつ人物を想定。

①　機械学習、データマイニング、数学、統計学などの専門家

②　分析専門のコンサルタント

③　大規模データ処理やＢＩ（Business Intelligence）／ＢＡ（Business Analytics）のシステムデザイン及び実装ができる専門家

2. 富士通（ヒアリング）

ヒアリングの概要（３／３）

☞　ビッグデータの活用は、単に企業の持っているデータを分析して経営指標に使えるというものとは異なり、非構造化データ等の色々なデータを分析して新サービスの立ち上げ・企画のためには、これまでのＢＩ等とは異なる数学的なノウハウや統計的な学力等を持ち、最終的にそれをビジネスやシステムデザインに結び付けることが出来る人材となるキュレーター（データサイエンティスト）の育成が重要。

☞　色々な機器にセンサーを付けてネットワーク経由でデータを集めると、ネットワークコストが膨大となるため、いかにコストを下げるかが重要。現時点では、通信ネットワークを共通化するために、他者と協力しての　標準化等に取り組んではいない。

☞　複数の企業や地方自治体等のデータを掛け合わせることにより新たなサービスが可能になるため、それらのデータの扱い方や、ネットワークを通じてクラウドセンターで扱うにあたって、セキュリティ的な問題が発生。

☞　個人情報の取扱いに関する法的な問題として、ＥＵのデータ保護指令により、ＥＵにおける個人のデータを日本に持ち込むことができないことと、医療データ等のセンシティブな個人情報を取扱うにあたっての扱い方について、緩和が重要。

☞　データの利活用を促進し、新しいサービスにつなげていくためには、個人情報等のデータについて、それらをどう使っていけばいいのかというコンセンサスやガイドラインのようなものを行政や企業等も含めて策定することが重要。

2．富士通（追加質問・回答）

1. 秘匿化について

　基盤技術の一つとして「秘匿化」を挙げていらっしゃるが、具体的にはどのような技術に取り組まれているか。
　また、そのような技術的なサービス提供に伴う制度上の不安や不明瞭な事項が、事業推進上の阻害要因になることはないか？

【富士通(株)からの回答】
　利用技術：データの暗号化や、独自のフェデレーション技術を利用しております。

課題　<個人情報において>：
　個人を特定できない匿名化技術が求められる。

- 秘匿以外に、個人情報を、個人が特定できないデータとして、統計や計算などに活用する場合、可逆不可である匿名化が求められる（健康情報など）。
- また、匿名化であっても、複数のデータ（事業所・年齢）で個人が特定化できるものは、個人データの扱いとなる。
- したがって、個人データを匿名化して扱うには、個人が特定できない、Kー匿名化を実施する必要がある。しかし、Kー匿名化は、すべての組み合わせをチェックする必要があるので、計算に時間がかかり、ビックデータを扱うには課題となる。
- 上記制約により、健康保険向けのデータ分析サービスなどを実施するためには、業務委託契約が必要であり、広く蓄積データを活用できない課題がある。

2.前項の関連事項として「店内カメラ画像解析」について

　店内カメラの映像をマーケティング目的で解析する場合、映像利用について、何がしかの許諾を取ることを行われているか？それとも運用上の工夫によって解決しているか？（顔が移り難い角度から撮影するなど。）
　もしも実験レベルであり、実用していないようであれば、今後、本件に関して想定される課題と解決施策、という観点でご教示ください。

【富士通(株)からの回答】

- 現在は実験を開始したばかりの段階であり、実用化しておりません。運用上の工夫において、個人を特定しにくいような映像とするように検討しております。
- 想定（期待）するデータ分析結果が出るのか自体が課題であります。また、運用等の課題については、実験を通じて抽出し、解決方策を検討していく予定です。

2．富士通（追加質問・回答）

3.ベテランスタッフの知恵共有について

農業従事者や医師の知恵をシステムに落としこむことを目指すとされていましたが、実証を進められる際に、阻害要因となる事項がなんであったかご教示ください。（他人のノウハウに学ぶ文化がそもそもない、低リテラシ等）

【富士通(株)からの回答】

＜食・農業分野＞

1. データを収集するための生産者が入れやすい簡易なUIが必要。
   現状：GPSやセンサを使ってなるべく自動化。タッチパネルUIなども活用。
   課題：音声や画像（動画含む）による入力方法があっても、検索やディクテーション能力がソフトウェア側に不足しており、蓄積したデータを活用する方法がまだ乏しいと考えております。
2. ユーザ（特に生産者）へのデータ活用の価値説明がしきれていない。
   現状：無理やり先にデータを入れてもらうか、アウトプットの事例を提示して理解してもらう等、工夫して取り組んでおります。経営面でどのようなデータが必要なのかをテンプレート化する等の取組みも必要と考えております。
   課題：全国規模での生産者のリクエストを把握出来ておらず、システム汎用化が可能なのかが不透明な部分がある。
3. 地域に散在する農業技術・知識を基盤として整備する必要がある。
   現状：限られた地域、品目において属人的なスキルを用いて整備を実施中。
   課題：当社単独の投資では面的展開は困難であり、自治体や国による支援を期待致します。

＜在宅医療＞

1. 医師やスタッフのノウハウの共有について
   - 医療、介護、看護、薬局などでのルールの違いや、情報リテラシーの違いなどが存在するため、多職種間で情報を共有する際には、入力や表示等においても工夫が必要となっております。
   - 院長が高齢化しているケース等では、電子カルテや在宅医療クラウドなどのICT化への理解を得ることにハードルが高い場合が多いのが現状です。
2. 今後乗り越える課題
   - 個人情報の取扱い
   - 医療データの共有化の壁
   - 多職種間での標準指標のオーソライズ
   - 医療系クラウドに対する規制緩和
   - 在宅医師の負担軽減

2. 富士通（追加質問・回答）

4.キュレーターについて

キュレーターとして活躍されている方はどういうバックグラウンドの方であるのか？（例：研究所で基礎研究に従事していた方を現場に選抜、など）
現状の規模（人数）はどの程度いらっしゃるのか？また、育成プログラムや、選抜基準はどのようにされているのか？
今後、どのように整備されようとしているのか？あるいは、外部委託も含めて検討されているのか？）

【富士通(株)からの回答】

- バックグラウンドは、主に以下の3点です。
  ① 機械学習、データマイニング、数学、統計学などの専門家
  ② 分析専門のコンサルタント
  ③ 大規模データ処理やBI/BAのシステムデザイン及び実装ができる専門家

- 現状は二桁の人数規模であり、今後3年以内に三桁にしたいと考えております。

- 選抜基準としては、保有するスキルセットや経験（上記のバックグラウンド）等を考慮しております。現在、研究機関他との共同研究や委託研究を実施している他、専門部隊が作ったロジックやツール等を広く展開していくため、分析のセミプロを育成するプログラムの策定等の施策を検討しております。

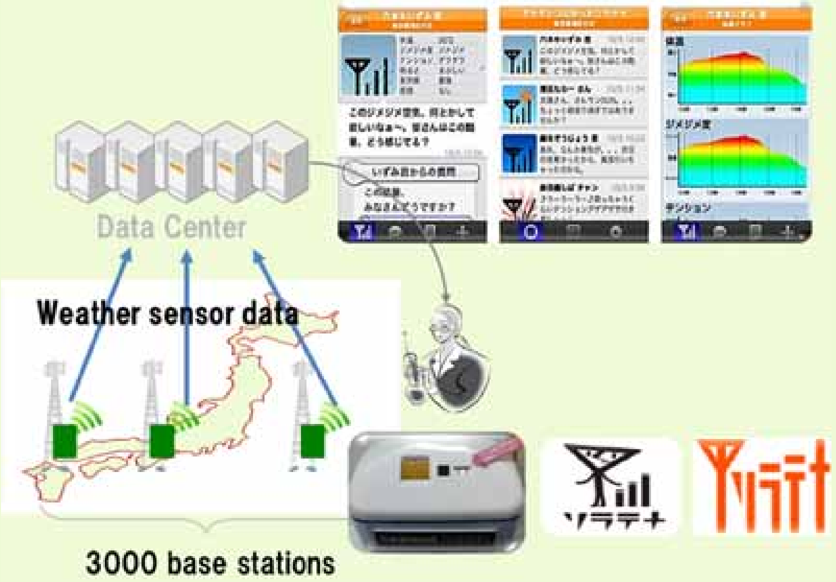
3．ＫＤＤＩ（取組事例）

取組事例の概要

- ａｕの全基地局の約１割に相当する全国3,000箇所の基地局に気象センサーを設置（アメダスは全国1,300箇所）。気温、気圧、湿度、日照、紫外線等の天候データをリアルタイム、ピンポイント、高密度で毎分測定。

- ａｕ基地局のセンサーから集めた気象データにコミュニケーションの要素を組み込んだ「ソラテナ」サービスを、ウェザーニューズ社と共同で2011年６月に提供開始。

- 擬人化したアンテナ「ソラテナ」が、気象状況を体感情報として届けるほか、今の気持ちをつぶやいたり、ユーザからコメントをもらったりすることで、新感覚のコミュニケーションサービスを提供。



- 通信会社として各種の通信ログ（通話の場合はＣＤＲ（Call Detail Record）、メールの場合はMailログ）を取得しており、発信者、着信者、開始時刻、終了時刻等の情報が含まれる。

- 通信会社には、法的対応（捜査機関への対応等）のために、通信ログを一定期間保存する義務がある。しかし、通信ログの情報は通信の秘密及び個人情報に該当するものであることから、現在、課金、法的対応及び通信品質の改善に用途を限定して使用。

3．ＫＤＤＩ(ヒアリング)

ヒアリングの概要

☞　ＣＤＲやMailログの活用により、ユーザのソーシャルネットワーク、基地局レベルでの位置、おおまかな行動パターン等を推定することが技術的に可能。これにより、リコメンド等のパーソナライズサービス、各産業等へのマーケティングサービス、人口動態分析等の社会的な貢献等が、サービスとして想定可能。ただ、現状では、通信の秘密の観点から、通信ログの利用は課金、法的対応及び通信品質改善という自社内の目的に限定。

☞　他方、欧州や米国では、次のように、匿名化という条件の下でＣＤＲを解析に利用可能。
①　SwissComは、ＣＤＲからユーザ間の人間関係を抽出し、ユーザをプロファイリングすることにより、個人情報の誤りや不正利用を検出。
②　AT&Tは、ニューヨークやロサンゼルスの近郊都市を対象として、ＣＤＲの活用により通勤経路や人口動態を推定。行政側もこの情報の利用を希望。ある自治体は、竜巻等の災害時にAT&Tからデータをもらい、竜巻の予測進路上のユーザに警報や竜巻の通過後の安否確認を行うサービスを提供。この点、米国では、通信法において、電気通信事業者は、匿名情報（集合情報）を利用することができると規定され、これを踏まえ、AT&Tのプライバシーポリシーにおいて、匿名情報・集計情報の利用が明記。

☞　日本においても、災害対応等の公共目的のために、匿名性を確保した通信ログによるサービスという形で貢献できるのではないかと期待。匿名化については、研究所においても、K-匿名化の研究が相当プライオリティが高い。ただ、明確に納得性のあるところまでというのは、まだ非常に高いハードルがあり、鋭意検討中。行政側に何か期待するというよりは、まず技術側で深掘りする段階。

☞　最近非常に気にしているのはＯＴＴ（Over The Top）の台頭。Google、Facebook等のＯＴＴは、従来、通信会社しか知り得なかった通信の秘密に該当する情報（ユーザのソーシャルネットワーク、正確な位置、行動パターン等）について、スマートフォン上のアプリを通じてユーザの同意の下で着実に蓄積。また、検索システム等において、ユーザの趣味・嗜好の情報も蓄積。サービスの実例として、Googleは収集した各ユーザの位置情報の移動状況から車に乗っているユーザを抽出し、交通状況を推定。

☞　ユーザと明示的な契約（Opt-in）を結び、統計的に匿名化を確約でき、通信の秘密に関する問題に対処できれば、通信ログの利用も可能とすることが必要。ヒトの位置データや行動データに、モノのデータを組み合わせてデータマイニングすることにより、新たな商品・サービスを提供したり、エネルギーマネジメント・都市計画等に活用が可能。

3．ＫＤＤＩ（追加質問・回答）

追加質問

ＫＤＤＩ様資料の７ページの冒頭のところに「米国や欧州では、ＣＤＲも、匿名化すれば解析に利用可能」と記載があるが、ご紹介頂いた事例の国においては、特に匿名化してあるデータであるならば、利用することを制限するような法律とか規制はないという理解でよいか。

【ＫＤＤＩ（株）からの回答】

とりいそぎ米国における匿名化情報の扱いを回答致します。（欧州等について正確な情報を確認次第、別途回答いたします。）

①　米国では、通信法－合衆国法典第４７編（２２２条）［※次ページ以降参照］において、電気通信事業者は、匿名情報（集合情報）を利用することができると規定しています。

以上を踏まえ、ＡＴ＆Ｔのプライバシーポリシーにおいて、匿名情報・集計情報を利用することが明記されています。

②　日本では、通信事業者が、通信の秘密に係る情報を匿名化（集合化）し、通信当事者の識別性を完全に喪失させたうえで二次利用するような形態であっても、電気通信事業法上、通信の秘密の侵害に当たると解釈される可能性があるため、当社では現状このような二次利用を行っていません。

ＩＣＴ利活用の活性化の観点から、我が国においても、通信事業者による匿名化情報の利活用について各方面から議論されるべきと考えます。

3．ＫＤＤＩ（追加質問・回答）

SEC. 222.［47U.S.C. 222］ PRIVACY OF CUSTOMER INFORMATION.

(a)～(b)　＜略＞
(c) CONFIDENTIALITY OF CUSTOMER PROPRIETARY NETWORK INFORMATION
　(1) PRIVACY REQUIREMENTS FOR TELECOMMUNICATIONS CARRIERS　＜略＞
　(2) DISCLOSURE ON REQUEST BY CUSTOMERS　＜略＞
　(3) AGGREGATE CUSTOMER INFORMATION - A telecommunications carrier that receives or obtains customer proprietary network information by virtue of its provision of a telecommunications service may use, disclose, or permit access to aggregate customer information other than for the purposes described in paragraph (1). A local exchange carrier may use, disclose, or permit access to aggregate customer information other than for purposes described in paragraph (1) only if it provides such aggregate information to other carriers or persons on reasonable and nondiscriminatory terms and conditions upon reasonable request therefore.
(d)～(e)　＜略＞

第222条[47U.S.C. 222]　　顧客情報のプライバシー
(a)～(b)　＜略＞
(c)　顧客に関する専属的な網情報（CPNI）の秘密性
　(1)　電気通信事業者のためのプライバシー要件　＜略＞
　(2)　顧客の要請に基づく開示　＜略＞
　(3)　集計顧客情報　－　電気通信サービスを提供することによって顧客に関する専属的な網情報を受領し又は取得する電気通信事業者は、(1)に規定する目的以外に集計顧客情報を利用し、開示し又はこれに対するアクセスを認めることができる。地域電気通信事業者は、当該情報の提供に対する妥当な要請に対して、妥当で非差別的な条件によって、他の通信事業者又は人に当該集計情報を提供する場合にのみ、(1)に規定する目的以外に当該情報を利用し、開示し又はこれに対するアクセスを認めることができる。
(d)～(e)　＜略＞

SEC. 222. [47U.S.C. 222] PRIVACY OF CUSTOMER INFORMATION.

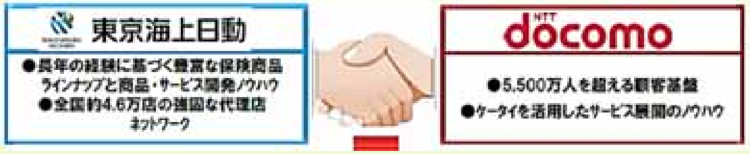
(f) DEFINITIONS- As used in this section:
(1) CUSTOMER PROPRIETARY NETWORK INFORMATION
- The term "customer proprietary network information" means--
(A) information that relates to the quantity, technical configuration, type, destination, and amount of use of a telecommunications service subscribed to by any customer of a telecommunications carrier, and that is made available to the carrier by the customer solely by virtue of the carrier-customer relationship; and
(B) information contained in the bills pertaining to telephone exchange service or telephone toll service received by a customer of a carrier; except that such term does not include subscriber list information.
(2) AGGREGATE INFORMATION- The term "aggregate customer information" means collective data that relates to a group or category of services or customers, from which individual customer identities and characteristics have been removed.
(3) SUBSCRIBER LIST INFORMATION　＜略＞

第222条[47U.S.C. 222]　顧客情報のプライバシー
(f)　定義　－　本条において用いる語辞は、次の定義に従う。
(1)　顧客に関する専属的な網情報　－　「顧客に関する専属的な網情報」の語辞は、次のものを意味する。
(A)　電気通信事業者の顧客が加入する電気通信サービスの数量、技術構成、種類、宛先及び利用総額に関する情報で、通信事業者と顧客の関係を理由としてのみ顧客が通信事業者に利用させるもの。
(B)　通信事業者の顧客が区域内電話サービス又は長距離電話サービスに関して受領した請求書に記載された情報。
(2)　集計情報　－　「集計顧客情報」の語辞は、サービス又は顧客の区分又は種類に関する集計データで、個人顧客の身元及び特徴が除去されているものを意味する。
(3)　加入者リスト情報　＜略＞

4. 東京海上日動火災保険/ＮＴＴドコモ（取組事例）

取組事例の概要（1／2）

- 2010年３月、ＮＴＴドコモと包括的業務提携。携帯電話を活用した新たなコンセプトの保険商品として、同年４月に「ドコモワンタイム保険」を販売開始し、種目は、①スポーツ・レジャー保険、②ゴルファー保険、③海外・国内旅行保険、④１日自動車保険。



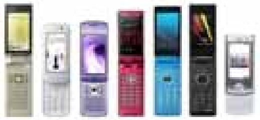
- 以上の取組による効果としては、例えば、次のとおり。
  - ①　ドコモワンタイム保険加入者の年齢層を見ると30代・40代が中心であるが、80代の人も少数ながら携帯電話を使って加入しており、このような年齢層のユーザにもアピールできている。
  - ②　ドコモワンタイム保険は、加入時期が正確にわかる。ゴルファー保険購入のタイミングを見ると、プレイ日前の深夜とプレイ日の早朝によく売れており、保険代理店を経由する必要なく、顧客のニーズに対応が可能となり、新たなマーケットを開拓。

- また、2011年７月からは「ドコモ医療保険」も販売し、種目は、①ベーシックプラン、②三大疾病重視プラン、③女性疾病重視プラン。健康支援の特典があり、ウォーキングやフードノート（普段の食事メニューの登録）等によりドコモポイントを入手（最大150ポイント/月）。ドコモポイントは、100ポイント＝100円でドコモグッズや修理サービスに使用可能。

- 携帯電話については、いつでも・どこでも・簡単に加入できる上、購入時刻の特定、加入者の本人確認、加入者の意思確認及び保険料の徴収が容易であり、新しい保険加入のスタイルとして期待。

4．東京海上日動火災保険／ＮＴＴドコモ（取組事例）

取組事例の概要（２／２）



- これらの保険の特徴は、次のとおり。

①いつどこででも、保険が必要になった日に加入可能

→　空港に向かう電車の中でふと不安になった時に海外旅行傷害保険に加入する、ゴルフのプレー前に調子が良いと感じたら、ホールインワンを達成した場合に備えてゴルファー保険に加入するなど。

②加入から支払いまで、携帯電話で全て手続可能

→　本人確認については、例えば、ゴルファー保険の場合、暗証番号、生年月日、プレー日の３ヶ所のみ入力。また、保険料（人が気軽に払える金額として、300円～）は電話料金と同時支払いが可能。

③いつどこでどのような保険が必要かを自動配信

→　ユーザの希望の下で、携帯電話のオートＧＰＳ機能により、場所に応じた保険をお勧め。例えば、空港に着いたら旅行保険、ゴルフ場に着いたらゴルファー保険、ユーザの生活圏から50km以上離れるとレジャー保険をお勧めなど。なお、通りがかっただけでお勧め表示を出さないための仕組みとして、目的地に５～10分留まった時に初めて表示されたり、また、ユーザの行動パターンを学習し、毎日行く勤務先等ではお勧めが表示されない。

4. 東京海上日動火災保険／ＮＴＴドコモ（ヒアリング）

ヒアリングの概要（１／２）

☞　今後の課題としては、次のとおり。

①　ワンタイム保険自体の認知度の向上（リピート率は高いが、認知度が低い現状）
②　販売で得られたデータを分析することによる、より良いサービスの提供
③　ユーザのニーズを把握した上での商品ラインアップの拡充
④　オートＧＰＳ（自動お勧め機能）の便利さの周知及び当該機能の利用者の増加
⑤　スマートフォンに対応するためのシステム開発（本年３月開始目途）

☞　どこからどこまでがスキー場や空港なのかという地図情報については、ＮＴＴドコモがゼンリン社との様々な提携により活用。

☞　ユーザの位置情報自体はＮＴＴドコモから提供を受けておらず、わかるのは保険加入があったという事実のみであり、どこのゴルフ場でゴルファー保険に加入したのかという情報は受けていない。センシティブな情報は通信キャリア内にとどめておくことが今後も必要。

4．東京海上日動火災保険/ＮＴＴドコモ（ヒアリング）

ヒアリングの概要（２／２）

☞ ドコモ医療保険については、手ごろな保険料でいつでもどこからでも加入できるという加入側のメリットがあり、また、保険会社にとっては、既存の代理店とは競合しない顧客層の加入者が増えるというメリットがある。

☞ 携帯電話のモデルにおいても約款は省略等せず全体版を載せているが、できる限りスクロールを少なくし、同意のポイントを幾つか絞ることにより、ユーザの作業を減らすよう工夫。

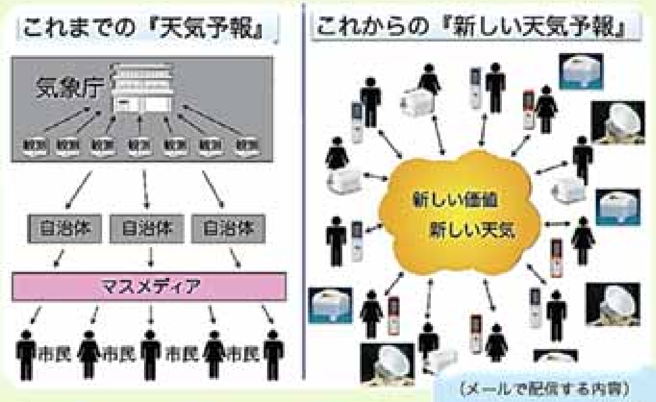
☞ 情報の提供を受けることのリスクもあるが、情報の種類によっては、特に、例えば、体の情報等については入手したいところもあるが、１つ１つの場合について、リスクとビジネスチャンスを考えながら対応する予定。

5．ウェザーニューズ（取組事例）

取組事例の概要（１／２）

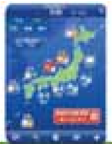
- アメダス等の観測器による観測データの解析と放送メディア等による一斉配信のみならず、「皆で参加（Join）して、情報を共有（Share）」というコンセプトにより、ウェザーリポーター制という有料の会員制度の下で、人間の五感による感測データを解析し、各利用者に最適なメディアによる配信を行う天気予報サービスを提供。現在、会員数は約28万人で、1日あたり5,000通のリポートが収集（台風時等には１日３万件あたりの報告）。

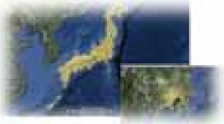
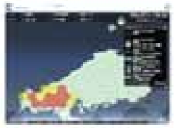
- 以上による効果としては、例えば、次のとおり。
  ① 60分先までの予報精度について、気象庁の「降水ナウキャスト」と比較して、約15%改善。また、降水量1mm以下の雨（人がポツポツと肌で感じるが観測器では観測できない雨）について、40%以上見逃すアメダスと比較して、会員の報告により把握可能
  ② 関東で雪が降った日（2011年２月）において、ウェザーリポート数が1日あたり1万４千通となり、アメダスによる観測が約17Km間隔であるのに対し、リアルタイムの降雪状況の把握が可能。
  ③ 予測が困難なゲリラ雷雨について、隊員数が3万人を超える「ゲリラ雷雨防衛隊」により、登録者が20万人の「ゲリラ雷雨メール」利用者へ80%以上の確立で事前にメール送信が可能。



5．ウェザーニューズ（取組事例）

取組事例の概要（２／２）



- 東日本大震災での取組としては、災害の記憶を記録する「減災レポート」を提供し、Google Map上において、津波、ライフライン停止、ガソリンスタンド・スーパー・コンビニ、電話・メール・携帯電話、帰宅困難、道路ひび割れ、液状化、火災等に関するレポートを表示し、各データのKMLファイルを提供。また、震災後においては、神奈川県、千葉市、名古屋市において、同様の減災プロジェクトを実施中。

- その他の取組としては、例えば、次のとおり。
  ①　個人等向けビジネス（BtoS）として、現在、160万人の有料会員に及ぶモバイル及びインターネットサービスや、放送メディア等を活用したサービスを提供。
  ②　企業向けビジネス（BtoB）として、企業等2,000者及び地方自治体1,000団体を顧客とし、航海気象、航空気象、道路気象、防災気象、農業気象、通信気象、電力気象、報道気象、放送気象、スポーツ気象、健康気象、生活気象、植物気象等のサービスを提供。
  ③　気象情報について、観測ネットワークとしては、アメダスの約1,300カ所（降雨量）の2倍以上の規模で、全国約3,000局の携帯電話基地局（au）に設置した、気温、気圧、雨量、風速、二酸化炭素量、紫外線量、感雨等の気象センサーにより、観測情報を収集。
  ④　局地的で、かつ短時間で変化する気象現象の捕捉及びその他気象現象に対する具体的対応策の提供を目的として、6秒に1回の観測、数百メートルメッシュの空間解像度、そして、車による機動的な移動観測が可能な「WITHレーダー」を全国75箇所に設置し、従来の観測ネットワークでは捉えられない、ゲリラ雷雨、竜巻や突風等の実況の把握等を実施。
  ⑤　欧州から日本へ通常の半分の距離で行けることにより、燃料の大幅な削減とCO2排出量の削減が可能となる北極海航路の支援や二酸化炭素のモニタリングのために打ち上げた独自の「WNI衛星」による観測データを活用。

5．ウェザーニューズ（ヒアリング）

ヒアリングの概要（１／２）

☞　気象データについては、利用用途が多岐にわたり、それぞれの用途に応じて、正確な精度が必要なものやそうではないもの等がある中で、現在、気象業務法においては、検定に合格した気象観測器で観測された気象データ以外のデータについてはインターネット等による公開が不可能。例えば、携帯電話基地局に設置している「ソラテナ」という観測器による気象データについては、そのまま公開することができないため、気象庁の観測データや検定に合格した観測器によるデータ等からマップ状にした実況解析値に、１kmメッシュ毎に落とす等の一定の処理をした形で公開。

☞　観測された気象データについて、実況ではなく、予測や予報に反映する場合についても、その観測器は検定に合格することが必要。例えば、ソラテナの気象データについては、予報には活用されていない。なお、検定に合格する観測器は１台あたり数十万円かかるが、エアコン等に搭載されている観測器は数百円程度であり、また、検定にあたっての審査費用や再検査にあたり観測器を取り外して持ち込むことが必要。

☞　気象データは軍事データとした扱われた時代があり、その後、昭和27年に気象業務法が制定されたが、その時点においては、気象庁自らが観測することは想定されておらず、検定に合格した観測器による気象データを予報に活用可能とする観点から基準が統一された。しかし、その後、気象庁において、アメダスにより自らの観測網を構築し、そのままそれ以前の制度が残っているというのが歴史的な経緯。

☞　気象業務法においては、ウェザーリポーターのような人間による報告データを予報に活用することは可能。他方、地震計については、同法において検定基準がなく、設置し、観測データを活用しても問題ない状況であり、観測するデータが同法の検定基準の対象かどうかで判断されている状況。

☞　海外等では、民間事業者がつくった観測器について、その売買が国により制限されておらず、市場原理に委ねられつつ、どれぐらいの観測精度が担保されているか等の情報が開示。そして、開示情報に虚偽等がある場合には罰則の対象になるというのが、観測器に係るルールとして海外等では一般的。

5．ウェザーニューズ (ヒアリング)

ヒアリングの概要（２／２）

☞　企業向けビジネス（ＢtoＢ）は主に問題解決型で、顧客のテーマに対しどう解決していくのかという形である一方、個人等向けビジネス（ＢtoＣ）は感動共有型であり、その点で「Join and Share」というコンセプトはＢtoＣに特化。

☞　基本的にウェザーリポーター等の会員から送られた情報については、ウェザーニューズに帰属し、そのライセンスにおいて、様々なメディアで使用することが前提。

☞　収集等した気象データのフリーライド又は流用のおそれについては、それらのデータは無料でインターネット等で公開しているため、いい形で他に使用される場合には協定を締結した上で使われることを考えており、厳しく対応していないのが現状。

☞　企業向けビジネス（ＢtoＢ）において、提供する予報は外れる場合については、基本的に、契約書において免責している。年間を通して、ウェザーニューズのサービスの提供を受ける場合において、費用対効果が最終的な料金の上限に関係するため、一つ一つの正誤ではない形で、契約と紳士協定により対応しているのが現状。

☞　気象データの活用として、例えば、電力需要想定をより的確に反映するため、観測された現地の気温等のリアルタイムな変化等を予測に取り込み、予報モデルに上げることで電力利用の最適化が可能であり、今後、様々な形での高度利用が想定。

☞　データの正確性や観測器の精度の問題については、気象データのみならず、例えば、計量法等に関するものも想定されるため、情報化社会というより大きな視点から検討することが必要。

5．ウェザーニューズ（追加質問・回答）

追加質問

予想が外れた「ハズレレポート」の活用などにより精度向上を行なっているとのことでしたが、サービス開始後、経年的に見て予測精度は向上したのでしょうか。

【（株）ウェザーニューズからの回答】

予測精度は時間スケールが3時間以内の予測は格段に向上しました。

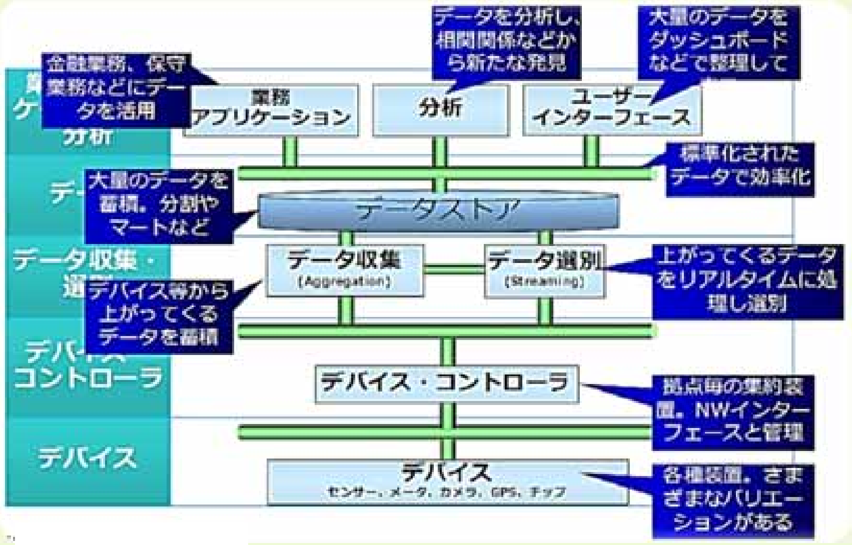
ゲリラ雷雨はこれまでは予測不能でしたが、９０％以上が捕捉できるようになりました。

これは、いまが分かればだいたい3時間以内の予測は分かる技術はあるものですが、こまかい現在が分かることが難しかったためです。無論、サポーターのリポートだけでなく、弊社では6秒間隔で観測する独自レーダーなどの技術革新もありますが、共通するのは今が分かる技術があがったことで未来がより分かるようになりました。

6．日本IBM（取組事例）

取組事例の概要（１／２）

- 世界規模での水不足、気候変動、食糧問題、医療システムの問題や大都市の交通渋滞等、様々な分野で問題が山積。他方、あらゆるものにRFIDタグやセンサー等が組み込まれて電子化され、相互接続が加速されて得られた情報のリアルタイム送受信が可能となり、また、収集された膨大なデータが瞬時に分析され、新たな洞察によりリスクや非効率を感知可能な環境が整備。このような状況を踏まえ、様々な問題解決に貢献する新しいコンセプトとして、地球をより賢く、スマートにしていく、「Smarter Planet」を推進。
- 「Smarter Planet」には、大量データの処理が可能な基盤、クラウド環境及び様々な標準化が必要であり、そのための基盤と構築サービスを提供。
- 「Smarter Planet」のアーキテクチャについては、デバイス、デバイスコントローラ、そして、データを収集・選別し、データをストアした上で、業務アプリケーションや分析につなげるという整理。特に、データを分析し、新しい洞察、知見につなげていくことが重要。
- ストリーム・コンピューティングの活用により、従来のデータベースのような技術ではなく、データを保存する前にリアルタイムに処理をし、その場でイベントに変換するリアルタイム・アナリティクスが重要。



6．日本IBM（取組事例）

取組事例の概要（２／２）

- 「ビッグデータ」技術の導入事例としては、例えば、次のとおり。
  - ①　米国政府による「Pacific Northwest Smart Grid」プロジェクトでは、ワシントン州等の５州における6万個の顧客メータを対象として、通信プロトコルの開発やデータの標準化等のため、全グリッドにおいて、どれだけ電気が使われ、今後の需要がどの程度見込めるかを全体でリアルタイムに管理し、風力等の再生可能エネルギーの使用促進等を実施。
  - ②　ブラジルのリオデジャネイロにおいて、交通や犯罪等について、情報ダッシュボードという形で様々なデータの一元管理が可能なシステムの導入による「Intelligent Operation Center」が設置。

- 収集された様々なデータの一元管理により、例えば、地図上にイベントの発生を表示する等、日常的な監視や緊急時・災害状況の把握を通じた災害対策や日常的な情報提供が可能。

- 大量のデータや非常に短い間隔で生成されるデータ等について、従来技術ではコストメリット的に困難であったが、現在の技術では、それぞれのワークロードに適したシステムやミドルウエアの活用により、トータルのシステムコストを抑えた新たな分析・情報基盤の構築が可能。

- 「ビッグデータ」の活用については、①種類（Variety）として、数値データ等の構造化データの他、画像や映像等の非構造化データ、ツイッターのつぶやき等の準構造化データ、②生成速度（Velocity）として、センサーから秒間100件や1,000件で上がってくるデータや１時間に１回まとめて送られてくるデータ等、③量（Volume）として、蓄積量や秒間処理量等の３つの側面を捉え、人間では単純には見つけられない知見や洞察が生まれ、情報を共有することが重要。

6．日本IBM（ヒアリング）

ヒアリングの概要（1／2）

☞　様々な分野でsmarterになる取り組みが行われているが、その中でも、例えば、今後は、医療や交通等の分野に注目が来る可能性がある。

☞　データを収集する場合において、センサーデータからいろいろリアルタイムに吸い上げる仕組みが基盤としてないことも問題。

☞　データ処理におけるデータ選別について、選別する理由としては、ストレージ容量、処理速度やリアルタイム性の確保等様々であるが、全てのセンサーデータを必ずしも蓄積する必要がない場合には、興味がある、又は異常値だけを選別して保存する。他方、ディスク容量が許容される場合であれば、一定期間保存するために全部収集するという使い分けが出てくる。

☞　日本で活動する場合には、日本特有の法制度が存在するため、海外のケースがそのまま日本で使えるというわけではなく、また、日本国内においては、「ビッグデータ」がないという顧客が多く、海外と比べて意識が低いという側面もある。

☞　日本国内の企業において、データ活用の重要性に関する気付きについては、例えば、スモールスタートやスモールサクセスで、最初はデータを収集等するためのコストをかけずに、ありもののデータで知見を得るという成功体験を与える等のやり方もあるが、海外事例の内容やメリット等の説明やそれを導入する場合のメリット等の説明も重要。特に、分析という観点においては、なかなかその結果を予測することが困難であるため、分析により得られるバリューの訴求が重要。

6．日本IBM（ヒアリング）

ヒアリングの概要（２／２）

☞　企業内部のデータをいろいろ集め、他の企業外のデータと合わせ込むことにより、新しいことを見つけようとする取組みにおける障壁としては、そもそも企業内のデータが整備されていないという点が挙げられる。例えば、事業部毎の売上データ等について、紐付けが効かない場合や、ＩＤが結びつかない場合等のデータマネジメントの問題がある。

☞　スマート化しようとする場合に問題になるのは、医療や通信等の分野にも関係するが、データについて、どこからでも何でも集めて分析していいというわけでなく、セキュリティの問題がある。

☞　データの秘匿化が問題であり、データの中身そのものの暗号化と、暗号化した上でのＩＤの取扱いという２点について、日本では海外と異なっている点があり、それらを取扱うにあたっては、国内法に引っかかる可能性もあるため、慎重に考える必要がある。

☞　データの活用において、通信事業者については、通信の秘密の問題がある。また、企業内のデータという観点においても、個人を特定できるデータについては、例えば、ポイントカードの使用により、統計をとり、男女の別やどんな人がどんなものを購入したか等のデータを収集しており、このようなデータを取り扱う上では、ＰＰＤＭ（Privacy Preserving Data Mining）等の暗号化や秘匿化が必要。

7．プリファードインフラストラクチャー（取組事例）

取組事例の概要（１／２）

Jubatus　Sedue

- ビッグデータの活用については、2000年頃からグーグル内で動き出したデータ解析が黎明期。具体的に、ウェブデータ、音声データ、画像や動画データ等の多様な非構造データを活用するソフトウェア等の解析基盤技術が開発。この点、Map Reduce等の関連技術がグーグルにおいて実際に使用されてから研究コミュニティに論文が公開されるまでに３・４年かかり、そして、その後オープンソースとしての公開等実用化までにさらに５年程度かかっているため、ウェブの世界におけるビッグデータの活用については、グーグルやアマゾン以外の者においてはグーグル等から10年は遅れている。そのため、グーグル等により、その間、例えば、広告配信やデータセンター等について、特許やプラットフォーム等が押さえられているのが現状。

- ウェブサービス分野も大きな部分ではあるが、グーグル等と技術的･戦略的に対等に勝負できるところとして、ウェブサービス分野以外におけるビッグデータ解析の実用化、現状のシステムがまだ技術的にできない部分に注力することが重要。この点、これまでも大手ＳＮＳの秒間数千・数万のメッセージのリアルタイム分析や工場から上がってくるセンサーデータによる監視等に取り組んできたところ。具体的に勝負するところは、次のとおり。
  ①　ビッグデータの活用の可能性が未知数であるウェブ以外の他の産業領域における新しい技術･戦略･ビジネス構築が必要。
  ②　データの量よりも、データの多様化と分析の質･速さが重要であり、各分野それぞれの専門的知識や、予測･発見･整理等の深い分析やリアルタイム処理が必要。

これまで取り組んだデータ解析事例（抜粋）

解析概要	分析対象データ	分析内容
ECサイト分析（自動車販売）	アクセス履歴・アンケート・ログ・顧客情報	車の評判分析ランキング、SEM、マーケティングへ利用
分析プラットフォーム構築	複数Webサービス間を横断したユーザーの行動履歴	ユーザーの属性分析・コンバージョン最適化
大手SNSメッセージ分析	メッセージ内容が不適切かどうかを分析．秒間数千~1万	リアルタイム言語解析．人の判断基準を即時反映
ウェブアクセス履歴	第三者広告配信のログ（大手広告代理店）	IP・クッキーに紐づいたユーザーの属性分析．
図書名寄せ	国内外の数千万冊の図書・	著者・作品・出版物の名寄せ
ソーシャルマーケティング	TwitterなどSNSデータ、	人・製品・企業・政策などのSNSでの広がり分析

7．プリファードインフラストラクチャー（取組事例）

取組事例の概要（2／2）

- ＮＴＴ情報流通プラットフォーム研究所との共同開発による大規模データを解析するための処理基盤である「Jubatus」をオープンソースとして平成23年10月に一般公開。開発以降、３ヶ月毎に新機能を定期的に追加。現時点では、事例により学習した分類ルールに基づくデータの「分類」、与えられたデータと似ているデータを高速に見つけてくる「推薦」、入力値から連続値を推定する「回帰」、各種統計量を高速にリアルタイムに取得する「統計」というビッグデータ分析に必要な機能をサポート。同基盤の特徴としては、次のとおり。

①リアルタイム性

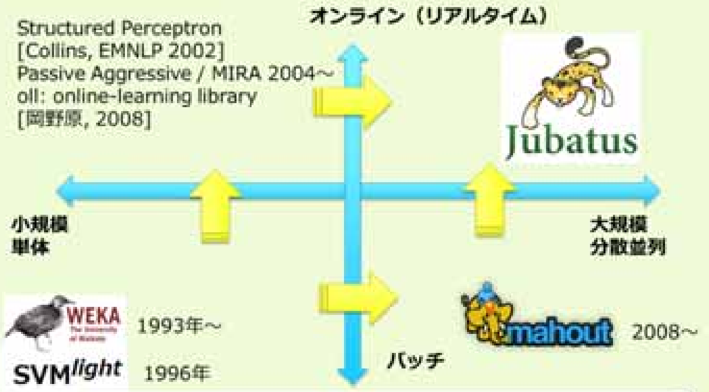
　データの到達と同時に処理を行い、リアルタイムに解析することが可能であり、極めて高い頻度で入力されるセンサーデータなどの情報から、即座に傾向を捉えてサービスにつなげていくことが可能。

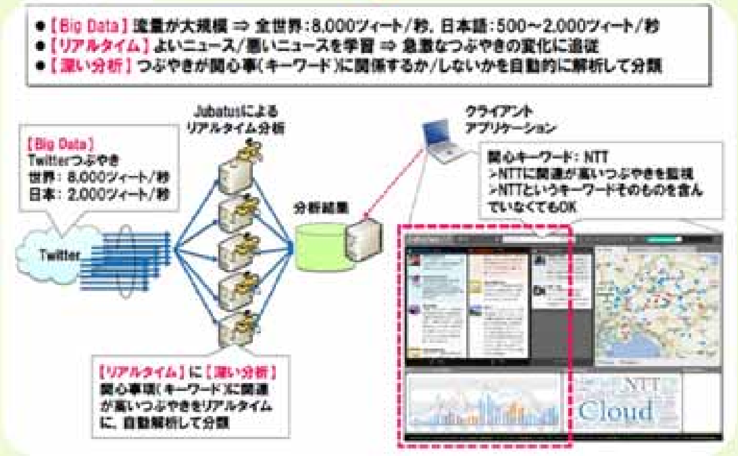
②大規模化

　より性能の高い計算ノードを利用するスケールアップ方式よりも価格面及び耐故障性の面で優れた、計算ノード数の増加により計算能力を確保するスケールアウト方式を採用。また、各計算ノードの独立処理による高い並列性の確保、低頻度での機械学習の結果の相互交換による効率の良い解析・分析が可能。

③　深い分析

　大量のデータから規則性や法則性を見つけ、過去の分析と将来の予測を行うための機械学習技術を採用。特に、データが次々とやってくる状況におけるオンライン学習技術を大規模分散環境で効率良く動作させるために拡張。また、生の非構造データを入力するだけで、機械学習による分析結果を得ることも可能。





7．プリファードインフラストラクチャー (ヒアリング)

ヒアリングの概要（1／2）

☞ ビッグデータの活用に関する今後の課題として、ビジネスをしていく上で感じたものは次のとおり。

① 言葉が先行し、貯めておいたデータに、何か技術を適用するとびっくりする知見が得られるというのは幻想であり、それに気づかずにインフラ投資に満足している例も多く存在。現在、ビッグデータを活用できているのはグーグル、フェイスブック、ツイッター、グリーやＤｅＮＡ等の一部のウェブサービスのみであり、この背景には、積極的に分析・サービスの改善に関わっているデータサイエンティストの存在が大きい。一方、その他の分野では、大きなデータの格納等のための大規模なハードウェア・案件の大規模化を目指したベンダー主導のプロジェクトが進んでおり、ノウハウ等が適切にロジックに組み込まれていないケースが多く、いわばビッグデータバブルの状態。

② 今後は、大規模データを扱えるインフラに加え、量や多様性の中から価値ある情報を見つけるための強力な分析ツール、データ分析のための継続的なノウハウの蓄積、適切なデータを取得できるようにする工夫の３点をバランスよく実施することが必要。そのためには、今あるデータのみではなく、継続的な取組が必要であり、ビジネスや対象領域への総合的な理解と、深い分析を行うための技術の両方の下で、仮説構築・検証・データ取得方法のＰＤＣＡサイクルを回し、その長期的な検証が必要。

③ 以上の取組のためには、既存のパラダイムや手法にとらわれていたら成功せず、今後爆発的に増加していくデータの種類が重要であり、統一されたフォーマットや手法で全てのデータを扱うアプローチよりも、整理可能な部分は整理し、整理できない部分は非構造化データのまま保持した上で、アドホックに秩序を見いだすというウェブの世界における仕組みが今後重要。また、量の増大よりも多様性の増加を生かし、様々な産業・組織がデータを組み合わせる価値を意識し、組織の構造や、日本という国全体でいろいろな産業にまたがってデータを分析できる仕組みが必要。

☞ 既にデータは収集されているが、活用できていないものとして、例えば、教育分野における学生の試験結果について、その後の就職等にも活かせる可能性があり。そのような動きは出始めている。

☞ 統計学等の博士号を有しているデータサイエンティストに加え、それぞれのデータ活用分野において、マネージャーレベルの人が統計等の深い知識を有している等、両方を知っていることが重要。

ヒアリングの概要（２／２）

☞　今後、ウェブを中心としてデータが多く収集等されているオンライン広告等のサイバーの領域から、多種多様なセンサー等からデータを収集等するフィジカルな領域にビッグデータ処理が進行。前者は新しい分野であるのに対し、後者は、今までの蓄積に対して保守的になる部分が大きいため、日本や米国においても、医療分野等の既存の産業に組み込むことは困難。今後は、よりＩＣＴ等のサイバー側の人材がフィジカルの分野に入り、データの取得等の観点から、センサーの設計等まで踏み込んでいくことが必要。

☞　ターゲティング広告等のウェブ分野では、新産業の創出やビジネス領域の拡大にはならない。今後は、携帯電話等の携帯端末により、人が今どこにいるのか、人の心拍数はどれくらいか等のバイタルデータも把握等することが可能になり、例えば、何か疾患が起こりそうな場合の事前予知等のＱＯＬを向上させる仕組み等、産業がかわるようなインパクトのある領域が重要。このような領域に投資しないと、マーケットセグメントを跨がるデータ融合も進まない。この点、社会インフラや医療分野に取り組んでいるところ。

☞　一部の人の属性や行動等がわかれば、全部のデータの収集が困難な場合であっても、高い精度で推定できるため、統計的な手法と組み合わせることにより、より詳細に今社会で何が起きているのかの分析は可能。データ収集の局面において、この部分を集めれば残りは十分に集まる等という統計の専門家からのアドバイス等のデータの収集とその後の統計解析の連携のための仕組みが必要。

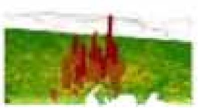
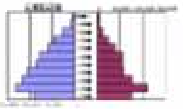
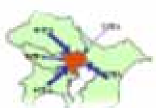
☞　データサイエンティストに必要な専門知識としては、統計、機械学習、データ構造、アルゴリズム、自然言語処理、画像処理等の幅広いものが必要。このような技術を持つ人は日本でもいるが、それらが全く違う分野で独立に動いているのが現状であり、連携してデータ解析ができるようになることが重要。

☞　従来はウェブにあるデータや英語や日本等の言語が多かったが、最近は、スマートフォン等の普及により、英語等を基本語としていた層以外の層もインターネットに参加してきた結果、例えば、ベトナム語やインドの地方言語等が量的に増加しており、また、テロ対策等でアメリカで注目されているソーシャルモニタリングにおいてはマイナー言語もサポートする必要があり、今後、統計的自然言語処理等による多言語解析も重要。

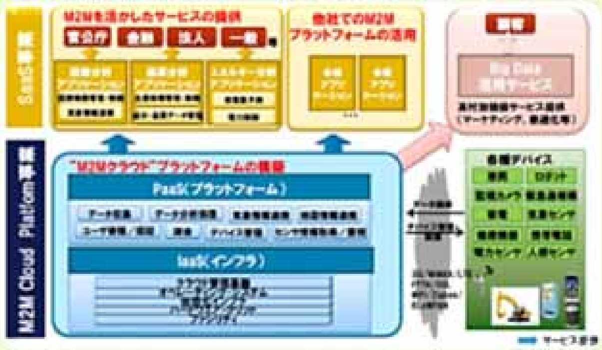
8．NTTグループ(取組事例)

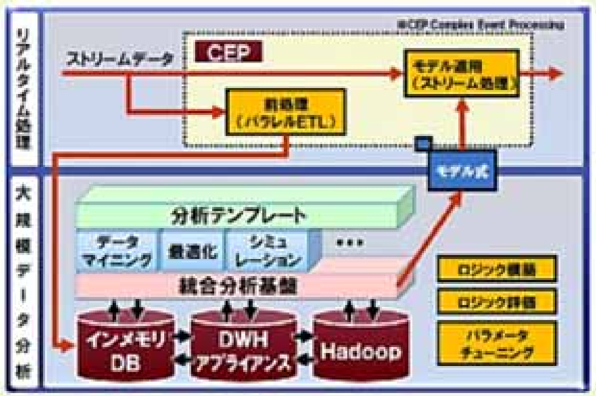
取組事例の概要（1／4）

NTTデータの総合クラウドサービス BizXaaS

- ＮＴＴデータにおける取組としては、次のとおり。
  - ①　平成23年10月に、M2Mクラウド推進室を設立し、Ｍ２Ｍクラウドの構築及びそれを活かした法人系の顧客への新規サービス提供（医療、農業、エネルギー等）を加速。
  - ②　Hadoopによるデータの分散処理等のBig Data基盤の構築、数理システム社への出資を通じた統計処理の専門家によるBusiness Analytics基盤の構築、そして、これまでのソリューション等のノウハウの蓄積等によるコンサルという分析テンプレートや分析APについて、基盤から分析までトータルにサポート。
  - ③　大規模リアルタイム・データ分析基盤においては、リアルタイム処理とは独立して大規模データ分析の仕組みを置き、データを蓄積・解析してモデルを策定し、ストリーム処理においてモデルを適用。
  - ④　「BICLAVIS」により、これまでやってきたものを次の9つに分類し、顧客の状況を見ながら、どこに向かうのがいいのかについて提案。
    1）評価・要因分析型（見える化）
    2）ターゲティング型（顧客のビジネスのやり方を改善）
    3）異常検出型（橋梁システム等の外れ値検出型、不正検出型）
    4）予兆発見型
    5）与信管理型
    6）予測・制御型（リスク・シミュレーション型、収益シミュレーション型、リスクヘッジ型、最適化型）
    7）コンテクスト・アウェアネス型（新たな知見の発見）
    8）マーチャンダイジング型
    9）プロセス・トレース型

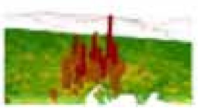
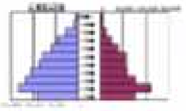
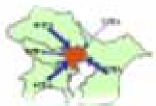




取組事例の概要（２／４）

株式会社 数理システム
www.msi.jp

NTTデータの総合クラウドサービス BizXaaS

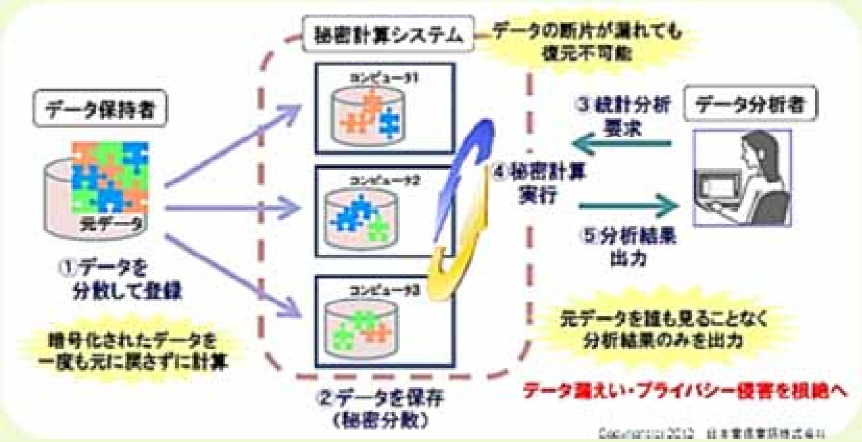




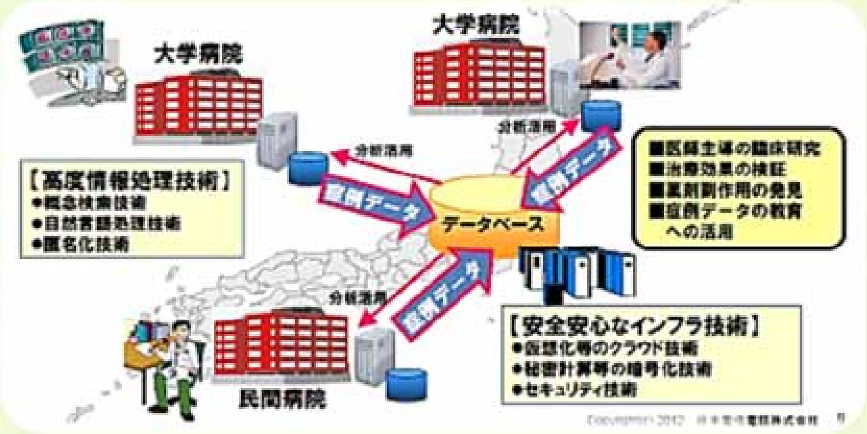
▷ ＮＴＴ研究所におけるビッグデータ技術開発の取組としては、次のとおり。

① ビッグデータのリアルタイム処理を可能にする大規模データ分析基盤技術として、ＮＴＴ情報流通プラットフォーム研究所とプリファードインフラストラクチャー（ＰＦＩ）社で「Jubatus」を共同開発し、平成23年10月にオープンソースとして公開。

② データを秘匿したまま計算する技術として、データを複数のコンピュータに意味のない断片に分散して保存（秘密分散）したまま算術演算やデータ操作演算（ソート等）を実現する秘密計算の効果的なアルゴリズムを開発。

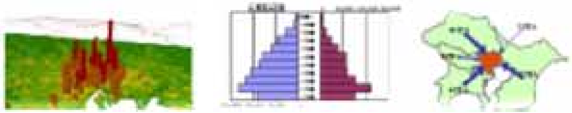
③ プライバシー保護、セキュリティ対策が必要な臨床研究データ（日本成人白血病治療共同研究グループ）を暗号化したままで統計分析を行う秘密計算技術を世界で初めて実証。今後、日本人の症例データに基づく、安心・安全な臨床研究データの環境を実現し、医療の質のさらなる向上へ貢献。





8．NTTドコモ(取組事例)

取組事例の概要（3／4）



- 携帯電話ネットワークは、携帯電話サービスの提供にあたって、運用データとして、顧客の住所、性別、年齢等や、携帯電話にいつでもどこでも着信したりメールを届けたりするための位置登録の情報等が必要。運用データについては、サービス提供のため以外にも、通常、ネットワークの設備設計として、通信量やトラフィックがどれくらい発生するか等の分析により、どの地域にどれくらいの基地局や交換機が必要か等のためにも利用。

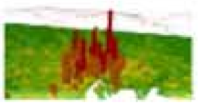
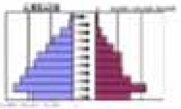
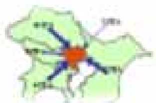
- 「モバイル空間統計」においては、携帯電話と基地局との定期的な交信により携帯電話の位置登録情報が運用データとして発生するという仕組みの下、顧客から申告された性別、年齢別、居住地別等の属性情報により位置登録情報を分類し、基地局毎に、例えば、男性が使っている携帯が何台、何十歳代の携帯が何台、どこの居住者が使用している携帯が何台という形で、地理的な人口分布（国勢調査等の全数調査ではなく、携帯電話をサンプル値として、基地局の各セルラー単位で人口を推計し、国勢調査と同様の500mメッシュに変換）、人口構成、移動人口という３つの空間統計を策定。

- 顧客のプライバシー保護のため、非識別化処理、集計処理、秘匿処理の３段階の処理を適切に実施。まず基本的な処理である集計処理として、基地局毎の携帯電話の台数を数えて、携帯電話の普及率をベースとして人口を推計。この前段においては、性別、年齢別、居住地別の人口で十分であり、個人を特定できる必要がないため、個人情報である運用データについての非識別化処理を実施。そして、集計処理の後段には、人口が少ない過疎地等において、他の情報との組合せにより個人が特定される可能性を排除するための秘匿処理を実施。

8．ＮＴＴドコモ（取組事例）

取組事例の概要（４／４）

NTTデータの総合クラウドサービス BizXaaS

▷ 研究開発段階の技術である「モバイル空間統計」の有用性については、専門家や自治体との共同研究を通じて、まちづくりや防災計画等の公共分野において、そもそもこのような人口統計に対するニーズがあるのか、ニーズがある場合に、そのニーズの要件にマッチしているかどうか、その要件にマッチしていない場合に、どのような改良が必要なのか等について、検証中。

目的	まちづくり分野で人口統計情報を用いる各種検討作業においてモバイル空間統計が実際に役立つ情報であるかどうかを専門家との共同研究を通じて検証すること
共同研究者	東京大学大学院　新領域創成科学研究科　社会文化環境学専攻 清家　剛　准教授（建築・都市計画研究者）
共同研究期間	2010年11月1日～
検証フィールド	千葉県柏市全域
内容	柏市役所の関係部署への課題のヒアリング結果をもとに決定した以下のケーススタディを実施 ①コミュニティバスを検討すべき場所の特定 ②土地利用区分毎の人口変動の把握 ③昼間人口に基づいた1人あたり公園面積の算出 ④中心市街地の来街者の特徴の把握

目的	防災計画分野の帰宅困難者対策の検討におけるモバイル空間統計の有用性を専門家や自治体との共同研究を通じて検証すること
共同研究者	工学院大学　建築学部　まちづくり学科 村上　正浩　准教授（防災・環境計画研究者）
共同研究期間	2010年11月22日～
検証フィールド	東京都全域（東京都 総務局総合防災部がオブザーバとして参加）
内容	30年以内に70%の確率で発生すると予測されているマグニチュード7.3クラスの首都直下地震を前提とした防災計画に必要となる以下のケーススタディを実施 ①東京都の帰宅困難者数の推計 ②東京都の徒歩帰宅者数の推計 ③帰宅困難者となる住民数の推計

8．ＮＴＴグループ/ＮＴＴドコモ（ヒアリング）

ヒアリングの概要

☞　ＮＴＴデータの「大規模リアルタイム・データ分析基盤」について、リアルタイム処理の入口にあるストリームデータの１つ１つは大きくないが、そのための通信の数が多くなる場合には、例えば、全てを３Ｇネットワークでやり、１セルに何万本・何万カ所となると、帯域的にはインパクトが出てくるかもしれないが、現時点では、そこまで大きくなっている事例はない。また、レイテンシ・遅延時間については、ストリームデータが次々と入ってくる場合において、通信の方がボトルネックにならない場合には、システムの方の条件次第。

☞　ＮＴＴデータの「ＢＩＣＬＡＶＩＳ」におけるコンテクスト・アウェアネス型について、例えば、コンビニ等のユーザの所へ、実際に経験者を派遣して、そのデータと物を診ながら個別にターゲットを決める作業を実施しており、経験・知見に頼って対応している状況。

☞　ＮＴＴドコモの「モバイル空間統計」で使用している位置登録情報は、CDR（Call Detail Record）とは違い、電気通信事業のガイドラインでもユーザの通信に係らない情報と分類され、統計処理をしても二次利用が不可というわけではない。

☞　「モバイル空間統計」において、位置登録情報の把握は基地局単位で１時間に１回程度把握し、非識別化処理等により何人そのエリアにいるかというのが分かるだけであるため、例えば、Ａ地点、Ｂ地点、Ｃ地点等という形で移動した人数は不明。携帯電話の位置登録の仕組み上、最新の所在情報しか基本的になく、その情報のみを活用して人口を推計するため、履歴的なデータは基本的になく、移動の履歴をベースとした移動人口は算出することができない。

☞　「モバイル空間統計」については、現時点で未だ研究開発中であり、そのニーズ、要件や改良等のために必要な共同研究のパートナーに提供しているという状況。この点、本統計を作成・提供する際に遵守すべき基本的事項について、自主的なガイドラインを作成して公開しているが、データの公開等の将来の在り方については未だ方針を決めていない。

8．ＮＴＴグループ（追加質問・回答）

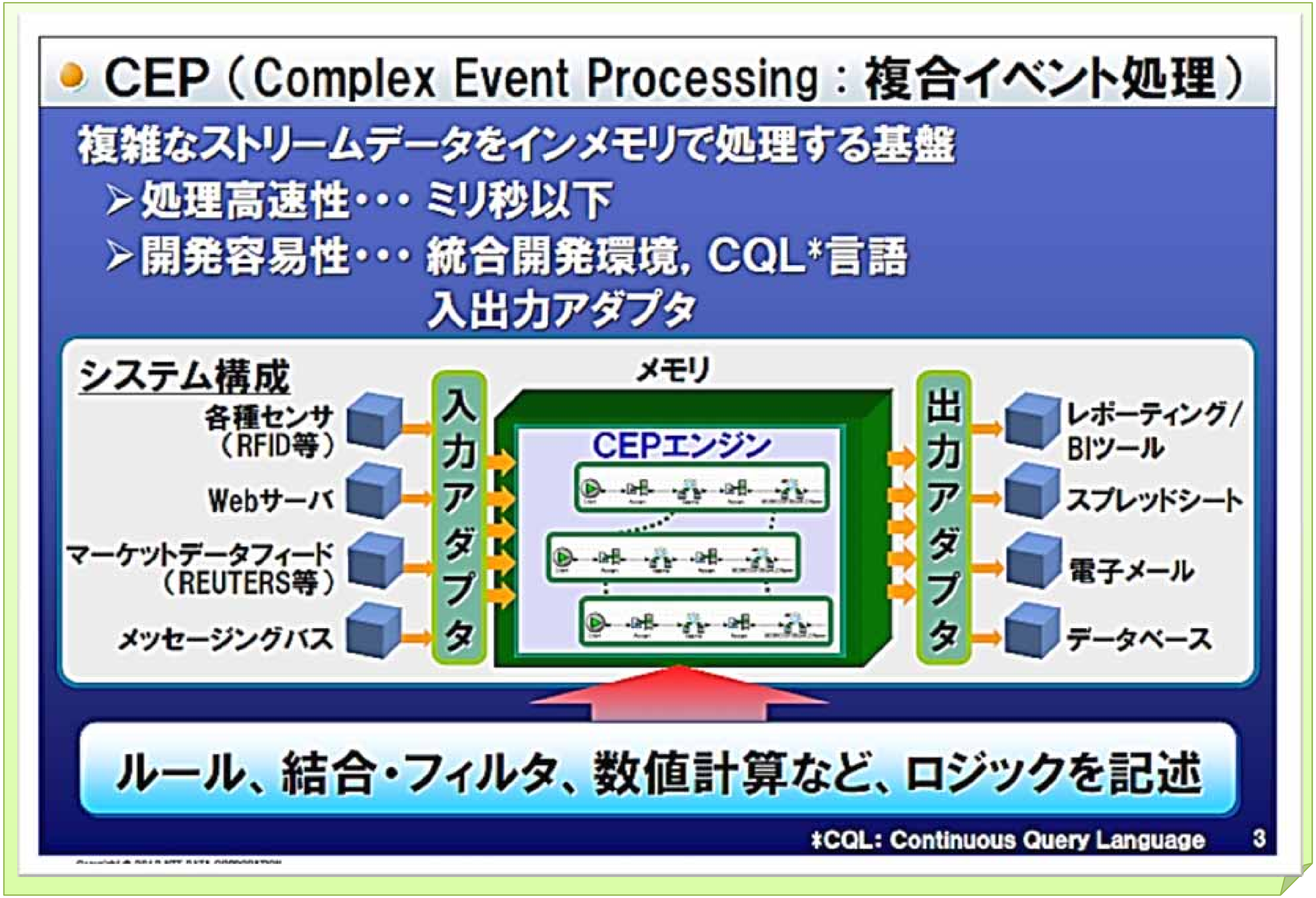
追加質問

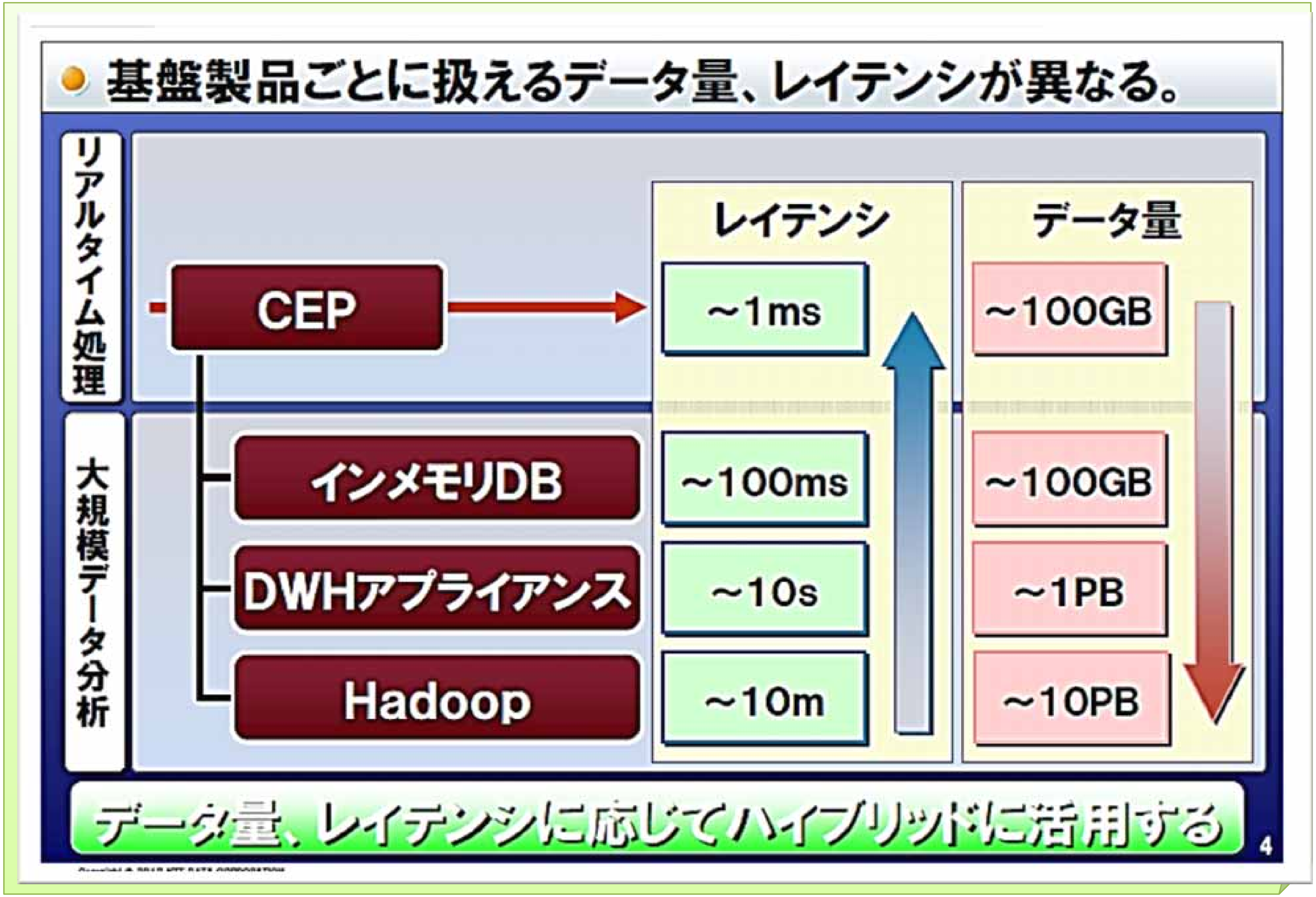
「大規模リアルタイム・データ分析基盤」におけるレイテンシ・遅延時間について、ストリームデータが次々と入ってくる場合、システムスペックの方はどのような条件となっているのか。

【日本電信電話(株)からの回答】

ＮＴＴデータでは、データ入力後、結果が出力されるまでのレイテンシを1ms以内に抑える必要のあるリアルタイム処理にＣＥＰを適用することを想定しています。

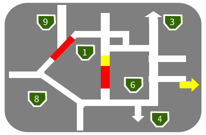
その際、ＣＥＰエンジンにおける工夫だけでなく、2ページ後の右側のデータ量で示していますように、処理の際に参照する辞書の大きさも制限することで、リアルタイム処理の実現します。

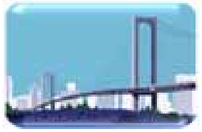

CEP（Complex Event Processing：複合イベント処理）
複雑なストリームデータをインメモリで処理する基盤
➢処理高速性・・・ ミリ秒以下
➢開発容易性・・・ 統合開発環境, CQL*言語
入出力アダプタ
システム構成
メモリ
各種センサ（RFID等）
Webサーバ
マーケットデータフィード（REUTERS等）
メッセージングバス
入力アダプタ
CEPエンジン
出力アダプタ
レポーティング/BIツール
スプレッドシート
電子メール
データベース
ルール、結合・フィルタ、数値計算など、ロジックを記述
*CQL: Continuous Query Language
3


基盤製品ごとに扱えるデータ量、レイテンシが異なる。
リアルタイム処理
大規模データ分析
レイテンシ
データ量
CEP
~1ms
~100GB
インメモリDB
~100ms
~100GB
DWHアプライアンス
~10s
~1PB
Hadoop
~10m
~10PB
データ量、レイテンシに応じてハイブリッドに活用する
4

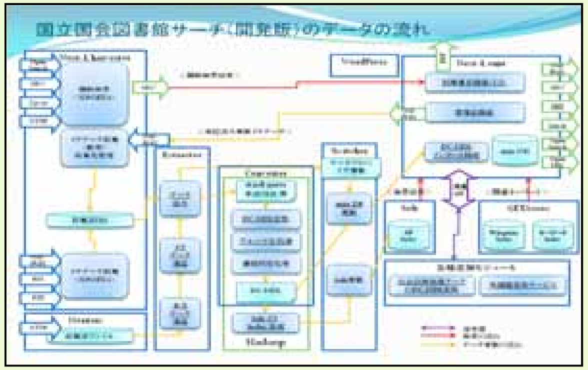
9．ＮＴＴデータ（取組事例）

取組事例の概要（1／2）

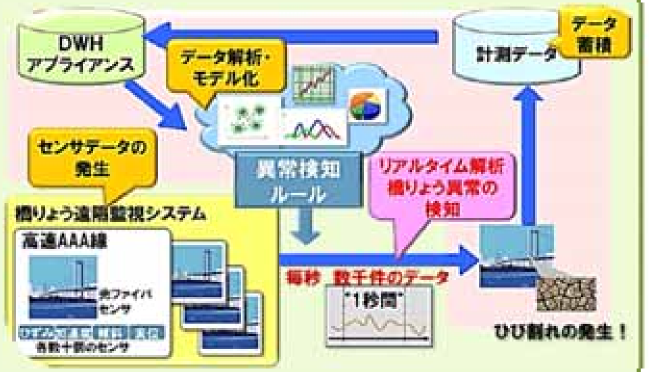




- ＮＴＴデータにおける取組事例としては、例えば、次のとおり。
  ①　国立国会図書館において、これまで蔵書リストは各図書館で管理され、国会図書館から全国の図書館等の検索は不可能であったが、Hadoopにより、バラバラなフォーマットや手入力による著者名等の蔵書情報をルール化し、蔵書データの同定することにより1つのリストとなる検索インデックスを作成し、国会図書館から全国の公立図書館・大学図書館・公文書館等の蔵書を横断的に検索が可能。



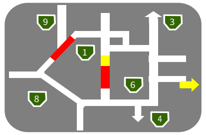
  ②　NTTデータの橋梁モニタリングシステム「BRIMOS®」において CEP（Complex Event Processing）を活用した大規模リアルタイム・データ分析に基づくプロアクティブ型BIの実証実験を実施。
  複数センサーの分析による総合的な橋梁異常検知を確認。
  実証実験では、橋梁に設置された多数のセンサーから発生するデータを対象とし、傾斜、振動、ひずみなどを算出する定型的な変換処理を行い、リアルタイムデータと過去の蓄積データも含めた複数の要素の分析に基づく橋梁の異常検知処理を対象に検証を実施。

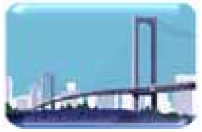


  ③　シンガポール科学技術研究庁では、空間的粒度、取得タイミング、データ形式が異なる複数ソースのデータを融合させ、時間別の渋滞について直感的に「見える化」する取組みを実施。また、降雨と交通情報の組合せにより、降雨の際の渋滞場所の情報を提供。

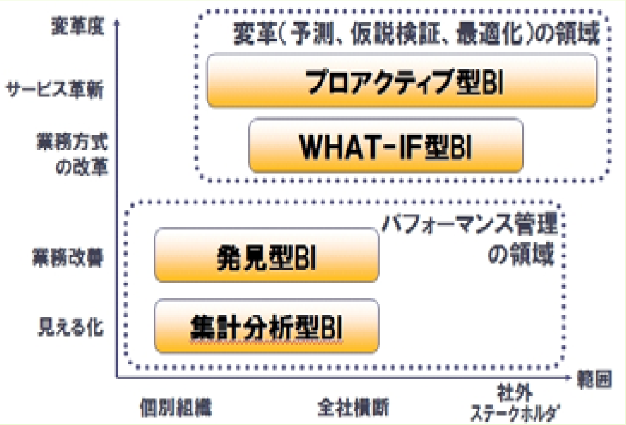
9．ＮＴＴデータ（取組事例）

取組事例の概要（２／２）

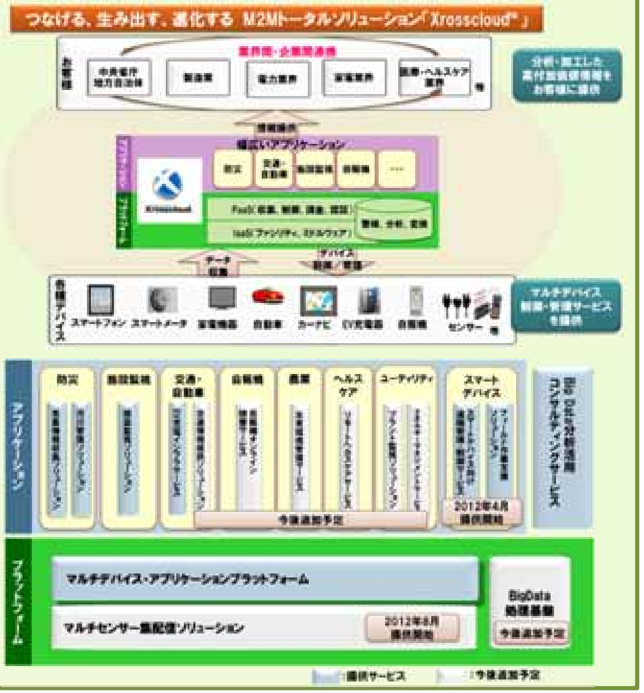




- ＮＴＴデータでは、次の4つのビジネスインテリジェンス（BI：企業内外に散在する膨大なデータを分析し、経営意思決定に活用する取組みの総称。）を提案。
  ①集計分析型BIとして、蓄積されたデータを様々な角度から集計・分析・見える化するもの。
  ②発見型BIとして、蓄積された膨大なデータから、隠れた関係性や規則性を発見するというデータマイニングの技術。
  ③WHAT-IF型BIとして、新しい業務方式をデザインし、その業務方式の変更に伴う効果をシミュレーションで試算するもの。
  ④プロアクティブ型BIとして、ユーザ行動を理解し、一歩先回りしてサービスや機能の提供をリコメンドし、不正が起こる前に不正検知するもの。



- ＮＴＴデータでは、M2Mシステムのクラウドプラットフォームからコンサルティングまでトータルで提供するM2Mトータルソリューション「Xrosscloud」シリーズを4月より提供開始。
  これまで提供している河川管理や気象情報の収集、電気自動車の充電インフラサービス等のアプリケーションをパッケージ化して提供するとともに、モノの情報を収集・管理し、その提供の方法等を管理するプラットフォームシステム基盤、ネットワーク、コンサルティング等をワンストップで提供。



ヒアリングの概要（1／5）

☞ ビッグデータについては、データに着目すると、従来ペタバイト以下位だったものが、ペタバイトからエクサバイトやゼッタバイトにどんどん増加するというもの。その理由は、例えば、センサーについて、設置可能数の増加や多様な種類のセンサーの設置の可能化、全ユーザーのデータの収集可能化、ウエブにおけるページビューからイベントトラッキングの可能化、１日１回の情報収集からイベント発生毎の収集等、非常にきめ細かく、また、リアルタイムに収集可能ということでデータ量が増加。

また、データ種別について、これまでは数値データ又はカラムが決まった構造化データだけとっていたのが、文書、メール、ツイッターのつぶやきやブログ、画像、静止画、動画等非構造化データも含まれる。

それから、データソースは、基本的に１カ所からのデータを今まで分析していたのが、複数データソースの情報を組み合わせて分析。

そして、処理に着目すると、例えば、今までデータ収集は離散的で、１日１回の収集等だったのが連続的なデータ収集になり、あるいは分析処理で言うと、バッチ的に処理していたのがリアルタイムの分析になるとか、また、データベースについて、リレーショナルデータベースでやっていたものがNoSQLのデータベースも活用するというような変化が生まれている。

以上のどれか１つとってビッグデータではなく、これらが複合的に組み合わさり、ビッグデータの時代と言われている。

☞ ビッグデータの活用は、次のように、より大規模なデータ、よりリアルタイムなデータ処理、よりデータの多様化へ進化していくと予測。

①個別利用と分析の時代（現在～２年）は、目的に応じたデータを収集し活用するプライベート型であり、数ペタバイトクラスのデータで、構造化（数値）データが基本となり、収集したものについて、ルールベースのリアルタイム分析とイベント検出に基づく制御を行う。

②共用利用と予測の時代（２・３年～）は、収集したデータを相互利用するコミュニティ型であり、数エクザバイトクラスの蓄積・分析として、構造化データのみならず、文書等の非構造化データとともにハイブリッドで分析し、モデルについても、過去の傾向に基づき動的にモデルを生成する方向へ移行。

③マーケットプレース化と総合・最適化の時代（５年～）は、データの流通を促し最適活用するマーケットプレイス型であり、数ゼッタバイトの蓄積・分析として、データフュージョンの下、モデルの自動更新とシミュレーションによる選択・最適化、リスク管理とともに、モデルに基づく予測と制御を行う。

ヒアリングの概要（２／５）

☞ ４つのＢＩについても、将来は、以下のように進化すると予測。
①集計分析型ＢＩについては、例えば、システムログ分析や四半期毎の売り上げ報告等から、渋滞と気象情報の組合せ等のデータフュージョンや必要なときに全件データを即座に集計、分析する方向で移行。
②発見型ＢＩについては、顧客行動分析等から、ＳＮＳインフルエンザ分析等、データ粒度を細かくすることにより組合せを増やす等へ移行。
③WHAT-IF型ＢＩについては、需要予測等の最適化と効果試算から、渋滞予測と信号機制御やデマンドレスポンス等、分析しリアルタイムに結果を戻して、その結果に基づいてまたトラフィック情報が変化し、それをすぐに収集して、情報収集してアクションするサイクルを回していく方向に移行。
④プロアクティブ型ＢＩについては、構造物の異常検出から、ライフログを使ったパーソナライズや解約の予兆発見等の方向に進展。

☞ 今後のビッグデータの活用には、高度な新たな処理基盤の導入と分析のデザインの両輪が必要であり、例えば、見える化して渋滞を予測して、それで信号や車線の規制を実施。この点、日本では、ＶＩＣＳやビーコンやプローブもあり、情報の収集は可能であるが、そのステークホルダーから情報を出してもらうのが大変な状況。また、海運ロジスティックスに関する将来例については、定量的に取得可能な固い情報（船の位置・大きさ、進路、速度、気象など）と、定性的なやわらかい情報（新聞、ツィッター、SNSなど）を組み合わせて分析し、各種の条件に基づく航路案を複数提示するなど、総合的な判断によりインシデントに対応するという俯瞰的な意思決定支援について研究中。

☞ Hadoopの元となった技術はGoogleによるところが多いが、ソース自身はオープンソースであるため、特にGoogleの技術であることについての懸念はない。Hadoopを開発しているコミュニティ（Apache Software Foundation）が買収されるというリスクはほとんどないと思われるが、万が一その場合でも、ApacheライセンスのOSSとして引き続き存在するため、ユーザや開発者にとっての影響は抑えられると考えている。また、ベンダー製品の場合はベンダーがサポートするが、オープンソースではサポート主体がないのが懸念であり、社内で技術者が実際に中まで見てチェックしているのが現状。

9．NTTデータ（ヒアリング）

ヒアリングの概要（３／５）

☞ ビッグデータビジネスに向けたハードルとして、以下の２つのハードルがあると考えている。
①ビッグデータ保有のハードルとして、現在はデータ保有者＝データ分析者であり、グーグルやアマゾン等が自分で行動履歴等を保有し、それを分析して各種サービスに役立てている。また、リアルなサービスについて、ログをとるのは困難であり、さらに、多様なデータを組み合わせて活用できる仕組みもこれからな状態。従って、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、著作権を踏まえ、まずは情報を使えるようにし、提供フォーマットやインターフェースの決定等、政府・自治体が保有するデータの一元的公開が必要。
②ビッグデータ分析のハードルとして、統計学者はいるがビジネス上の課題がわかっていない者が多く、ビジネス上の課題を理解し分析できる人材が不足。また、クラウドにより数千台のサーバを借りるのが楽になった反面、Hadoopのインストールに手間が発生している点が課題。そして、分析結果を実証する場がなく、データ活用イノベーションのコンテスト開催や特区での実証実験等、オープンイノベーションに向けた場の醸成が必要。

☞ 欧米では政府機関が保有している様々なセンサデータを公開し、民間に活用アイディアを競わせ、キラーアプリ／コンテンツを模索中。2009年５月21日に開設したアメリカのdata.govでは、約39万のデータセットがraw dataで公開され、APIや「意味（semantic）」を標準化し、データを民間等が自由に活用可能。また、地質環境汚染状況、肥満分布状況やオゾンの状況等、1,192の政府アプリと236の民間アプリが開発。さらに、2011年12月30日のGSAによる報告諸では、非常に人気が加速し、例えば、ダウンロード数も年間成長率270％で増加し、データセット数もかなり増加し、現在約39万となっている。そして、データやアプリを共同で作るコミュニティについて、オープンデータ、セマンティックウェブ、健康、法、エネルギー、海洋、教育の７分野で構築しており、コミュニティ内のデータコラボレーションで納税者への節税が5.5百万ドルだったと報告。

☞ イギリスでは、政府が保有する情報を民間のSNS等と連動させて新たな価値を生み出すことを目的とするPower of Informationプロジェクトとして、政府情報に関するオープンイノベーションの取組が実施され、2010年1月19日にはdata.gov.ukが開設。

ヒアリングの概要（４／５）

☞　データ収集におけるＩＣＴリソースの課題として、センサ等からの情報収集は、ネットワーク化するモノの多さに比して、効果創出が難しく、費用対効果のバランス点が総じて低いのが特徴。Last 1 foot問題として、数メートルではなく、100メートル以下位をどうやって飛ばすのかが問題であり、近いところにセンサーを一杯ばらまくと、そこでの無線の競合が発生するという課題がある。

☞　複数データの組合せによる新たな価値創造の例として、IC定期券によりどこからどこまで乗車したかがわかり、さらに、携帯GPSにより、優位性が出る程度の人数がそれらのマッチングを行えば、それらの者がどういうルートで動くのか、旅客の行動パターン、曜日や時間帯による変化、エリア属性、平日の朝にはこの辺りに人が集まる等がわかる。これにより、通勤ルートの状況を踏まえ、例えば、帰宅難民の予測や備蓄量の試算、エリアの混雑度による避難場所や誘導先の決定、また、都市計画への活用が可能。

☞　国が保有している情報について、著作権制度を変えるというわけではなく、不必要に著作権を主張せず皆が使えるようにすることが重要。

☞　個人情報については、法律を変えるという話ではなく、これ位までであれば使用可能という点について、実証の場等において、国が音頭をとり、得られるベネフィットとプライバシーとして失われるもののバランスの問題として、国民のコンセンサス等の醸成を図ることが重要。

☞　５年以上先におけるマーケットプレース化におけるデータが流通するためのデータの価値の定義について、その価値は使う人によって大分異なり、価格については、経済原理により買う側と提供する側の双方で決まると考えられる。また、流通するための条件としては、データ自身の提供ではなく、いわゆるメタデータとしてどういう形で定義していくかが考えられる。

☞　オープンソース化については、コミュニティでソフトを維持できるため、自分たちが作成したものは自分たちで維持する必要があるが、例えば、10人が使うと維持費用が10分の１になるという狙いがある。

9．NTTデータ(ヒアリング)

ヒアリングの概要（5／5）

☞ 日本における公共データのオープン化については、データがどこにあるかがわからず大変であり、また、フォーマットやデータをとってくる時の形式等、プロトコル自身が規定されていないのが現状であり、一元的に見せるような形にすること等が重要。

☞ 日本における公共データのオープン化については、形式がバラバラでも統合できることが今後の発展の中にあったとしても、データがどこにあるのかを明確化することのみならず、データの形式及びデータをアクセスする形式を合わせることが重要。

☞ 欧米ではオープンガバメントの推進、新産業創出の促進に加え、インドで共同開発することにより、インドにも通用するプラットフォームをつくるという戦略があり、今後、様々な分析手法やプラットフォームの在り方が、徐々にどこかにシェアが集まり標準化されていく可能生がある中においては、オープンイノベーションに向けた場づくり等、実証事業や、実際に収集したデータを用いて分析した結果をアクション等するところが重要であり、大変ではあるが、この分野は世界中でも今進み出したところであるため、これに取り組むことにより、ビッグデータの有り難みを皆が享受できるし、世の中も変わっていくと考えられる。

☞ 現在、ウェブ上のサービスにおいては、データを取得して分析した結果を活用する面で非常に進んでいるが、実社会における分析結果を活用したサービスの提供はこれからと考えられる。

☞ スマートシティやスマートコミュニティの実現は、海外においては大きな国ではなかなかやるのが困難であり、情報を収集してそれをアクチュエイトやアクションすることについては、ステークホルダーが多いと調整等技術以外の面でなかなか困難。他社の事例などでも多くはこれから発展するような先進途上国に対して実施しており、データ分析基盤を提供し、現地の国や研究機関等が分析するという形でやっている。

１０．インテル（ヒアリング）

ヒアリングの概要（１／３）

☞　ビッグデータの対応課題として考えられるのは、次のとおり。

①インフラの整備…今後５年先を見た場合、指数関数的にデータのトラフィックが増大する等、それに対応できるネットワーク基盤、コンピューティング・パワー及びストレージ容量が必要。日本のインフラは、世界でも進んでいると思われるが、これをもってしても、データ通信量が指数関数的に増加する場合に対応できるかどうかが問題であり、例えば、電力に伴う発熱等を考慮すると、100倍のデータ通信量のサポートのために、データセンターを今のアーキテクチャーやストレージシステムのまま100倍にできるかが課題。

②標準化…データには様々なフォーマットがあるため、１つのデータのある種の標準化や、大量のデータをメタデータ化するという前処理におけるホストコンピュータにあげるメタデータに関するフォーマット変換等のデータの処理や管理プロセスに対する標準化等が課題。また、標準化の対象等や、そのプロセスについて、インターネットポータル系の会社等によるデファクトスタンダード化となるのか、社会インフラ等のための活用におけるＡＰＩやインターフェース等に関する標準化委員会等における標準化となるのか等については、ちょうど議論が始まったところ。

③セキュリティ…データの収集、分析、活用におけるセキュリティ、個人情報、プライバシーが課題。

☞　データ量については、現在、ソーシャルメディアデータに代表される非構造化データが指数関数的に増えているが、2015年には、LTEやWiMAX等のWiFiホットスポットの普及等の無線を中心とするネットワークの整備により、Ｍ２Ｍ等のセンサーから上がってくるデータが非常に大きな加速度をもって増大し、ソーシャルメディアデータを追い越すと考えられる。

☞　ビッグデータについては、データセットが大きいか小さいか、そして、構造化データを扱うリレーショナル・データベース系か非構造化データを扱い非リレーショナル・データベース系かで分けると、今後非構造化データが大部分を占めてくることを考えると、大きなデータセットで非リレーショナル・データであるものをビッグデータとして位置づける傾向が多く、リアルタイム処理等の様々なサービスや手法等により、従来型の構造化データを中心とする伝統的なビジネスインテリジェンス等とのバランスをとりつつ活用することが重要。

10. インテル(ヒアリング)

ヒアリングの概要(2/3)

☞ ビッグデータの要素として考える必要があるのは、以下のとおり。

①量…ソーシャルメディアデータ等の非構造化データが急速に増加し、全体のデータの９割近くを占め、構造化データに比べ10倍から50倍、従来型のデータ・ウエアハウスに比べ10倍から100倍の速さで成長すると見込まれる。なお、さらに、センサーから来るデータを加えると更に大きな容量になると見込まれる。

②種類…データのフォーマットが様々で、大きさも異なっていることに加え、スキーマが無く又は弱く、さらに、文法や解釈方法が不揃いであるため、このような千差万別のものをどのように処理するのか等による分類が必要。

③価値…将来のトレンドやパターンの予測分析として、データ抽出技術、統計モデリングやグラフ・アルゴリズム等を駆使しながら、従来型ビジネス・インテリジェンスに対するより深い複雑な分析により新たな価値を見出せるかが重要。なお、この点については、利活用の目的を明確にし、投資額に見合うリターンを得られる方でインフラ投資できるかどうかが課題。

④スピード…バッチ型ではなく、リアルタイム分析として、例えば、ITSを含めた車交通の制御等においてはリアルタイム性が必要となるため、バッチ処理ではなく、リアルタイムで大量のデータを処理し、即時対応のサービスができるか否かが課題。

☞ ビッグデータを支える基盤としては、データセンター側のサーバ、ストレージ及びネットワークについては、ソフトウェアでフローをコントロールする技術や、Hadoopのみならずそれから派生した各種ストレージ技術も必要になると考えられる。また、データセンター集中型の構成だけではなく、エッジ・サーバやハブにおいて、ある程度の前処理を実施したり、リアルタイムのレスポンスにおいてデータセンターをティアリングするイメージでのシステムアーキテクチャが必要。さらに、ネットワークにつながるすべてのデバイスがデータソースとなり得るが、それらのデバイスに対して処理やサービスを返す基盤が必要。

☞ データのリアルタイム処理や分析等を行うエッジ・サーバの設置場所について、ユーザーサイド側にするのか、データセンター等の中央側になるのか等については、提供するサービスにより、サービス提供者が系全体の最適化等の観点から最適な場所を選択することになると考えられる。

10. インテル（ヒアリング）

ヒアリングの概要（3／3）

☞　現在は、利用目的に合わせ、データ形式がバラバラな状態で、アプリケーションセントリックなデータの収集、加工や処理等が中心となっているが、数年後は、M2M等の様々なところからデータが上がってくる中で、個々のソリューションがバーティカルに乱立するというよりも、ベースのところについては、標準化等の前段階の加工があり、その上に、個々のバーティカルなアプリケーションが乗る仕組みが必要となるという課題があって、その際、特にセキュリティやプライバシー等の水平分野に関わる共通化や標準化が必要になるのではないかと考えられる。

☞　計算のやり方については、これまでのジェネリックなやり方をベースにそれを工夫することで対応するのではなく、今後は、用途特化型の異なるアルゴリズムによるデータ分析技術が必要になってくると考えられる。

☞　中国最大のショッピングサイト「Taobao」では、店舗、ユーザーは無料であり、3.7億ユーザーと８億種類の取扱商品を抱え、2010年には１秒あたり800トランザクションという非常に大きなトランザクション量を扱い、2012年には約1500億ドルの年商となっている。そこで、Taobaoの売上げの80％が、ビッグデータの活用による個人に特化された広告サービスの展開で占められている。

☞　通信系のインフラ整備については、公共性が高いため、政府の役割も重要であり、また、データセンターのサービスについては、「.co.jp」からサービスが提供できる環境づくりも重要。

☞　まとめとしては、インターネット接続デバイスと非構造化データの爆発的な増加がビッグデータのニーズと活用を牽引しており、ビッグデータのトレンドを形成する４つの要素、そして、海外政府等におけるより良い社会づくり、効率的な社会インフラや安心で安全な生活への活用等のビッグデータの積極的活用も踏まえながら、ビッグデータ処理におけるインフラ整備や標準化等の新たなＩＣＴが必要。

国内におけるその他取組事例①

RECRUIT　HOT PEPPER

☞ 同社の運営サイト「ホットペッパー」について、利用者の利用履歴について1日で分析可能範囲が2週間分から2年間分に拡大

☞ 利用者への定期配信メールの開封率が1.6倍に上昇し、同サイトのページビューが急増

☞ 今後、旅行サイト「じゃらん」、中古車サイト「カーセンサー」等全てのサイトに同システムを導入予定

楽天　Infoseek 楽天

☞ ターゲティング広告について、リアルタイム性を加味したデータベースシステムにより、会員の多種多量に及ぶデータを集約・分析

☞ 数千万人の会員の属性、数千万点の商品購入履歴、各種サービスの利用履歴、会員ランキング、ポイント活用等のデータを日次等で分析し、顧客特性等に応じた広告等を配信

☞ 会員が2つ以上の他サービスを利用する割合が2007年の31.4%から2009年6月に38.2%に向上し、また、クリック率や購買率が数倍に上昇。

ANA

☞ 最新鋭機「B787」について、新たなロードコントロールシステム（LCS）により、数百人規模の旅客や預け荷物の数、搭載予定の貨物の重量等をネットワークを通じて瞬時に収集

☞ 運航の安全性と燃費向上の両面から機体の重心が最適な位置になるよう、全搭載物の配置を瞬時に割り付け

☞ 重量計算等の自動化により、離陸直前まで重心位置のシミュレーションを繰り返し、低燃費の追求が可能

国土交通省 関東地方整備局
Ministry of Land, Infrastructure, Transport and Tourism, Kanto Regional Development Bureau
東京都

☞ 平成24年2月に開通した「東京港臨海道路」の東京ゲートブリッジについて、多数のセンサーにより、橋のひずみや振動を常時検知し、橋の破損状況をデータとして把握可能

☞ 橋を通過する車両の重さを算出することにより、過積載を遠隔監視し、橋への負担や事故につながる車両の走行を防止することも可能

［出典：各社ウェブページ等を基に作成］

国内におけるその他取組事例②

紀の国 有田 早和果樹園 SOWA

- ☞ みかん栽培について、同果樹園内5カ所に設置した農業センサーにより、気温、湿度、土壌温度・水分、降雨量、日射量等の20種類のデータを収集
- ☞ 樹木5千本にIDを付与し、日々変化する樹木の育成状況や病害虫の発生状況をクラウド等により管理
- ☞ みかんの生育と水分吸収、降雨量と害虫発生量等の複数データから相関を分析し、次期生産活動に活用

KOMATSU　現場が見える！オフィスでできる車両管理 KOMTRAX

- ☞ 建機について、GPS（全地球測位システム）等により、稼働状況の遠隔監視等のため、位置情報や車両内ネットワークからの情報を収集
- ☞ サーバ側システムにおいて、車両から送信されたデータを蓄積し、インターネットを通じて、顧客や販売代理店に提供
- ☞ 自社機器の稼働データにより、建設需要が増大する地域の予測や、顧客におけるリアルタイムの稼働状況の把握による正確な与信確保等に活用

カブドットコム証券 わたしたちはMUFGです。 kabu com

- ☞ 投資情報について、分散処理ソフト「Hadoop」（ハドゥープ）を組み込んだシステム等により、ツイッター等のソーシャルメディア上のデータを分析
- ☞ 46者の対象銘柄について、1日あたり約900万行のソーシャルメディア上のデータを収集し約4万3千のキーワードで絞り込み、登場頻度と株価動向の相関関係等を分析
- ☞ 今後、対象銘柄の増加や分析精度の向上等の開発により、新たな投資情報サービスの提供が期待

徳島大学病院 Tokushima University Hospital

- ☞ EHR（電子健康記録）について、分散処理ソフト「Hadoop」やデータ管理ソフト「Cassandra」等を組み込んだシステムにより、医療機関等の診療等データを集積し、疾病を管理・分析
- ☞ 徳島大学病院や保健センター等のデータを同病院内のサーバに蓄積し、診療所20カ所の検査結果等を集約・分析
- ☞ 慢性疾患対策の観点から、継続的な健康情報の管理により、包括的な疾病予防管理サービスの提供が期待

［出典：各社ウェブページ等を基に作成］

海外における取組事例①

SANTA CRUZ

- 窃盗事件について、余震予測システムをベースとした犯罪予測システムにより、過去8年分及び日々更新される犯罪データを分析
- 同市内を500フィート（約152メートル）四方のエリアに区切り、最も犯罪が起きる可能性の高い地域トップ10が警察官に伝送
- 市民からの呼出が増加する一方で、警察官数が減少する現状におけるより効果的な人員配置や、窃盗事件に対する抑止効果が期待

Ford

- プラグインハイブリッド車について、いつ、どこまで、どのルートを運転したかに関するデータを収集し、分析
- 例えば、運転手が夜6時にエンジンを始動させる場合の目的地の予測や、交通渋滞等を回避できる最適ルートの提案等が可能
- 電気自動車専用ゾーンをハイブリッドカーで走行する場合における電気・ガソリン利用の適切な切替え等、燃費効率の向上が期待

UOIT

- 新生児集中治療について、予測分析システムにより、新生児に装着されたセンサーから送られてくるバイタル・データ（体温、心拍数、血圧値など16種類）をリアルタイムで収集・分析
- 予め登録された新生児の平均データから成る基準モデルと比較し、心肺停止や院内感染などの罹患リスクの存在を伝達
- 看護師による直接診断より6～24時間早く新生児の容態異常が検知可能となり、また、複数患者の同時観察により院内感染予測等に寄与

VISA

- クレジットカードの不正検知について、全会員の利用パターンを作成し、カードの利用・取引状況データを分析
- 数週間かかっていた全会員の利用パターンの作成が13分に短縮され、従来1ヶ月に1回だった不正検知用パターンの更新が1日に複数回可能となり、精度が向上

［出典：新聞記事等を基に作成］

海外における取組事例②

T-Mobile

☞ 解約者について、1ペタバイトを超える大規模なデータ分析システムにより、3,400万人の契約者データを分析

☞ 2週間分の分析により、一定の契約者グループの中の1人が解約し事業者を乗り換えた場合、同じグループの契約者が一般的な契約者と比較して7倍高い確率で乗り換えていることが判明

☞ 分析結果を踏まえ、解約する前に、契約者に対し、ピンポイントで優遇キャンペーン等を打つことが可能となり、経営が改善

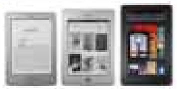
ERICSSON　MTN

☞ リアルタイムでの割引施策（Dynamic Discount Solution）を行ことで、トラフィックが集中する時間帯においては、料金を高くし、脆弱なインフラにおいても耐えられるよう通話を平準化

☞ ネットワーク容量の使用が過小である場合には、利用者が最大９５％までの割引を受けることが可能

☞ 通信事業者において、通話時間や通話回数の増加に加え、収益が改善し、契約者数が増加、解約者数も減少

amazonkindle
Read. Review. Remember.



☞ 電子書籍kindleにおいて、本人の同意により、電子書籍のハイライト情報をAmazonクラウド上に収集し活用

☞ どの電子書籍のどの部分がハイライトされているかという分析により、集計情報の見える化等による販売促進に活用

shopperception

☞ マイクロソフト社のモーションセンサであるKinectを活用し、スーパー等の小売分野において、買い物客の行動を分析

☞ 陳列棚におけるどの商品に客が手を伸ばす回数が多いのか等について、リアルタイムでセンシングし、ヒートマップによる見える化や最適に商品配置

☞ 商品を一旦手に取った後に陳列棚に戻す顧客に対し、デジタルサイネージ等の活用により、当該商品に関する広告をリアルタイムに提供

［出典：新聞記事等を基に作成］