

KDDI/KDDI研究所における ビッグデータ関連の取組

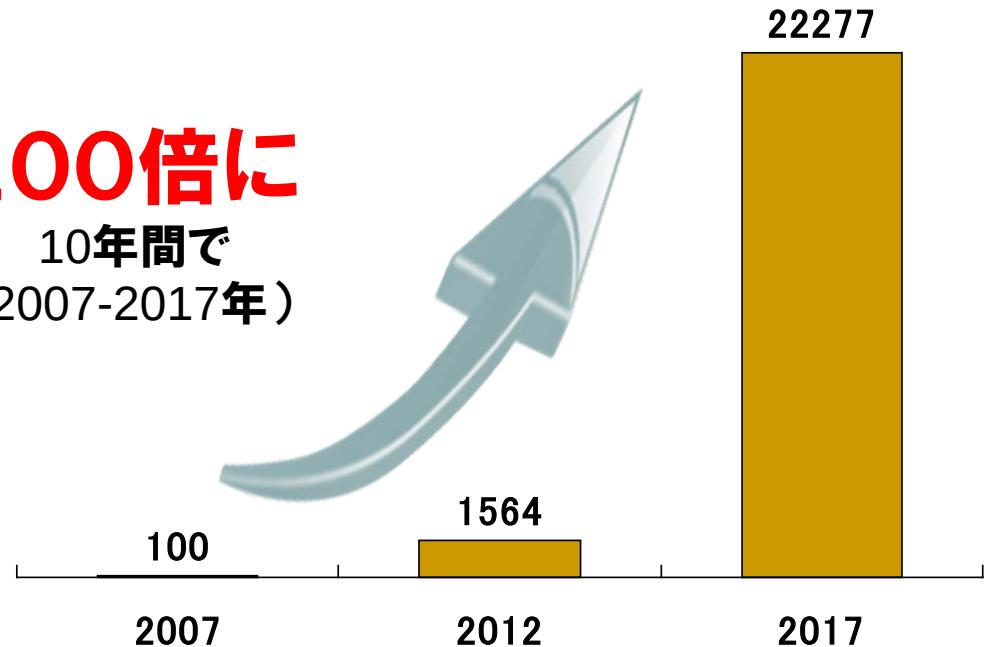
株式会社KDDI 研究所
グリーン・クラウド部門
杉山敬三

第一部 Big Data

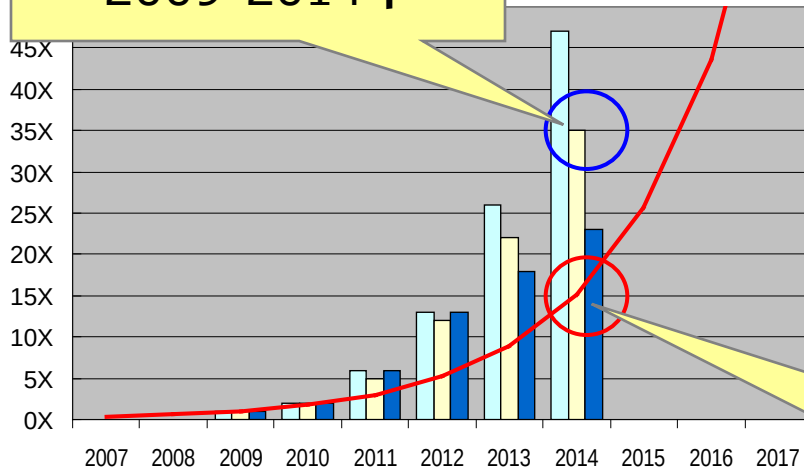
データトラフィックの急増（日本）

200倍に

10年間で
(2007-2017年)



北米では5年間で
35倍に
2009-2014年

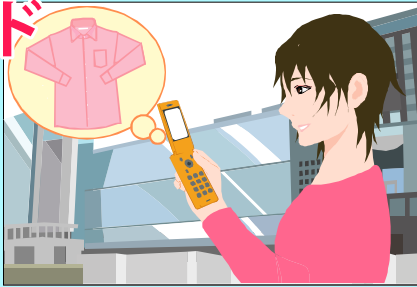


出典：情通審（携帯電話等周波数有効利用方策委員会、2008年12月）

日本では5年間で
15倍に
2009-2014年

ライフログの適用領域

レコメンド



何を買ったか？



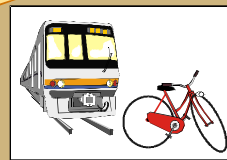
ログ



ヘルスケア



何に乗ったか？



何をしなければならないか？



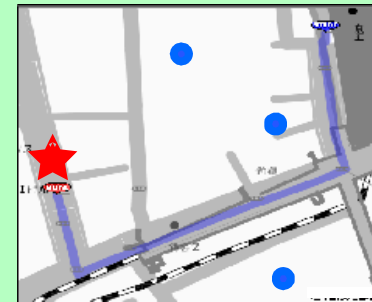
どこへ行ったか？



操作支援



備忘録



「Big Data」とは？

Wikipedia :

- 通常のデータベース管理ツールなどで取り扱う事が困難なほど
巨大な大きさのデータの集まり
(構造化データ+半構造化データ+非構造化データ)

例)



伝票



レシート

例)



Web画面

例)



ビデオ



音声



テキスト

日本IBM HPから :

例えば、

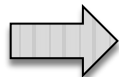
- ICタグなどのセンサー
- ソーシャル・メディアに掲載された投稿
- インターネット上に保存されたデジタル写真やビデオ
- オンライン購入の処理レコード
- 携帯電話のGPS信号

など、さまざまなソースから生成されるデータ。

「Big Data」とは？

これまでも、巨大なデータの分析は行われてきた：

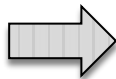
例) 衛星センシング情報、 ゲノム情報、 大規模小売店のPOSデータ



- 分析にコストがかかるため、分析対象は限定的
(要するにかなりの利益が見込める領域のみ)

近年

- 蓄積や分析のコストが急低下した
 - ハードディスクの大容量化に伴い蓄積ビット単価の急低下
 - Hadoopなどの新しい分析基盤技術の登場
- 通信（特に携帯電話）技術の発達によるデータの急増
 - ネット上のWebデータをクロールすれば誰でも「ビッグデータ」を持てる



- HW的には、ちょっと思い立てば分析できる状況に
「ビッグデータ」がはやり言葉に、、、
- でも、実際に分析技術を持つ専門家「データサイエンティスト」
が不足し、一般企業では活用が進まない状況に

Hadoop vs. グリッドコンピューティング

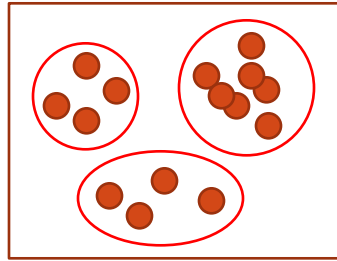
- 共通点： 多数のサーバでクラスタを構成し、スケールアウトにより性能アップ
- 相違点： ストレージに対する考え方

	ストレージ	特長	用途
Hadoop	ローカルHDD (分散)	読込性能	テキスト検索 データ集計
グリッド	SAN (共有)	演算性能	シミュレーション 科学計算

- Hadoopが提供するものは、信頼性の高い
 - 共有ストレージ (HDFS)
 - 分析システム (MapReduce)
- データを演算ノードに合わせて配置
 - Shared-Nothingアーキテクチャ
- MapReduce の得意分野
 - 書き出しは1度で、読み込みは何度も行われる用途
 - データ量が[1TB]以上ならアドバンテージが大きい
- 障害に対するリカバリを MapReduce が制御

データマイニングとは

データの中から隠れた役に立つ知識を取り出す技術



クラスタリング

類似したデータの
集まりは？

予測

次に発生し
そうなデータは？



相関分析

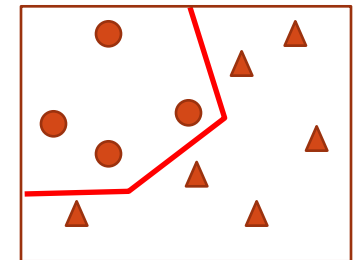
同時に発生し
やすいデータは？



AとBは一緒に買われやすい

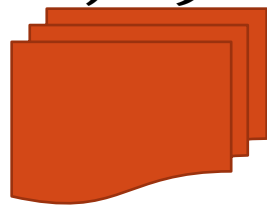
分類

データは
どのカテゴリ？



マイニングプロセスの例

データ



データ
変換

ノイズ
除去

変数
選択

モデル
選択

ルール
学習

モデル
学習

分類
予測
など

仮説検証の繰り返しにより、精度向上

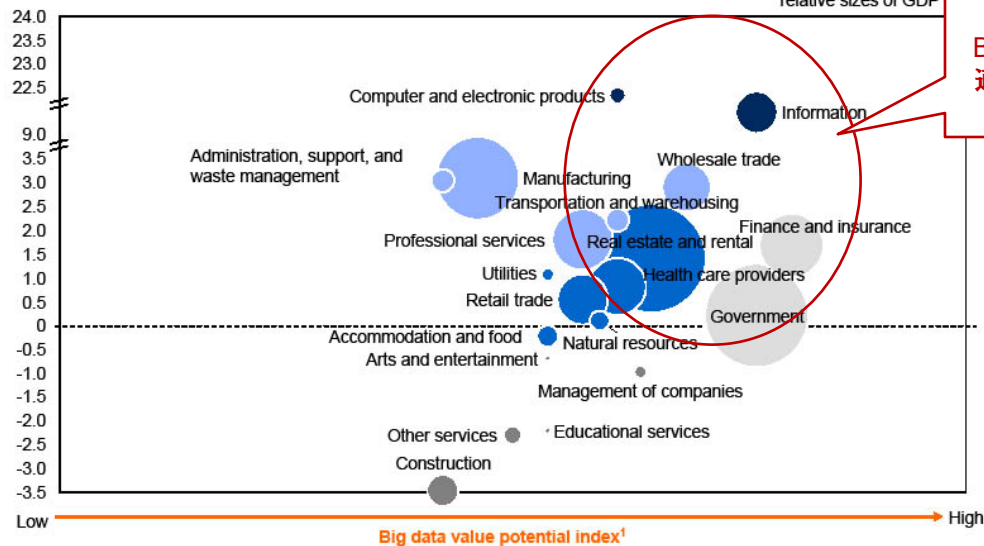
Big Dataの付加価値が高い産業

- マッキンゼーでは幾つかの産業を「Big Dataの付加価値が高い産業」として指摘している
 - すでに利用が進み、高い生産性成長のある産業：情報、コンピューター・エレクトロニクス（紺色）
 - 現在の生産性はやや低い、Big Dataによる付加価値が大きいと見られる産業：金融・保険、政府（灰色）
 - 国際競争がないため生産性成長が現在は低い、Big Dataの活用の付加価値が大きいと見られる産業：不動産、医療・ヘルスケア、小売（青色）
 - 国際競争の大きい産業は、Big Dataの活用有無にかかわらず、高い生産性をすでに実現している：製造、卸売、プロフェッショナル・サービス（水色）

Some sectors are positioned for greater gains from the use of big data

Historical productivity growth in the United States, 2000–08

%



*バブルの大きさは産業規模を示す

出典： McKinsey Global Insti

¹ See appendix for detailed definitions and metrics used for value potential index.
SOURCE: US Bureau of Labor Statistics; McKinsey Global Institute analysis

通信ログ活用の現状

通信会社は、

- 法的対応(捜査機関への対応など)のために、各種通信の利用ログ(通信ログ)を一定期間保存する義務がある
- 一方で、通信ログを通信会社としての本来の業務以外の目的に利用することは原則できない
(憲法で定める「通信の秘密」を守る義務がある)

KDDIにおいても現状は下記の目的に限定して利用

- 課金
- 法的対応
- 通信品質改善

諸外国の状況

欧米では、通信ログも匿名化すれば統計等に利用可能

米国（ベライゾン、AT&T等）

- 匿名化のもとで統計データに利用することを、プライバシーポリシーに明記
- マーケティング目的の場合については、内容によってはOpt-inを取ること、またユーザの拒否権(Opt-out)も明記

欧州（BT、ボーダフォン等）

- 匿名化のもとで統計データに利用することを、プライバシーポリシーに明記
- マーケティング目的については、Opt-outの権利を明記

第二部 M2M

M2Mの主な分野

ARPU

潜在市場規模(2013年予測)

出展: 矢野経済、TSR、富士経済、国土交通省

産業機器

187万

工作機械
建機など

乗用車
商用車

1,000万

プローブ
CAN
地図更新など

運行管理

26万

レンタカー
リースなど

エレベータ

監視

45万

緊急通報
遠隔監視

複写機

(MFP、LBP)

285万

トナー監視
故障検知

エネルギー
(電力・ガス・水道)

1,300万

遠隔検針
開閉・監視など

通信カーナビ

20万

車載ナビ
PNDなど

ハンディ
ターミナル

70万

情報照会
音声通話

自動販売機

258万

在庫管理
決済・認証など

警備・セキュリティ

300万

機械警備、位置監視
ホームセキュリティ
など

絶縁監視

42万

漏電監視
など

台数

M2M標準化領域

M2Mの相互接続性を高め、市場拡大を加速するために、M2Mのビジネス形態に応じた各レイヤでの標準化が重要。特に、サービスプラットフォーム、コアネットワーク、ゲートウェイを中心に活動。

レイヤ

標準化

M2Mアプリケーション

(公共、電力、医療、運輸、ガス、上下水道、工業、農林水産業、家電制御、次世代自動車など)



・産業セクタに特化した情報要素と情報流通方法の定義等。

→各種団体で実施

(例: HEMS Alliance, Continua Health Alliance 等)

M2Mサービスプラットフォーム



・全産業のM2M関連機能と情報要素および流通方法の抽象化 (→例: ETSI-TC-M2M、今後(は)oneM2Mで実施)

・M2Mデバイス管理 (→例: OMA DMとBBFで実施)

コアネットワーク

移動網
(GSM, 3G, LTE)

固定網
(ADSL, FTTH, CATV)

・既設網をM2Mへ適合させるための仕様の検討 (→例: 移動網では3GPP MTCで実施)

ゲートウェイ

ホーム/ビルディング
/ローカルネットワーク



エンタープライズ



・ローカルネットワーク (ホーム、工場等) に特化した情報流通の枠組みを定義

(→例: ETSI-TC-M2M、BBF、OMA DM、今後(は)oneM2Mでも実施)

無線

無線

有線

無線

有線

M2Mデバイス、M2Mエリア/センサーネットワーク

(3G、Z-Wave、ZigBee、UPNP等)

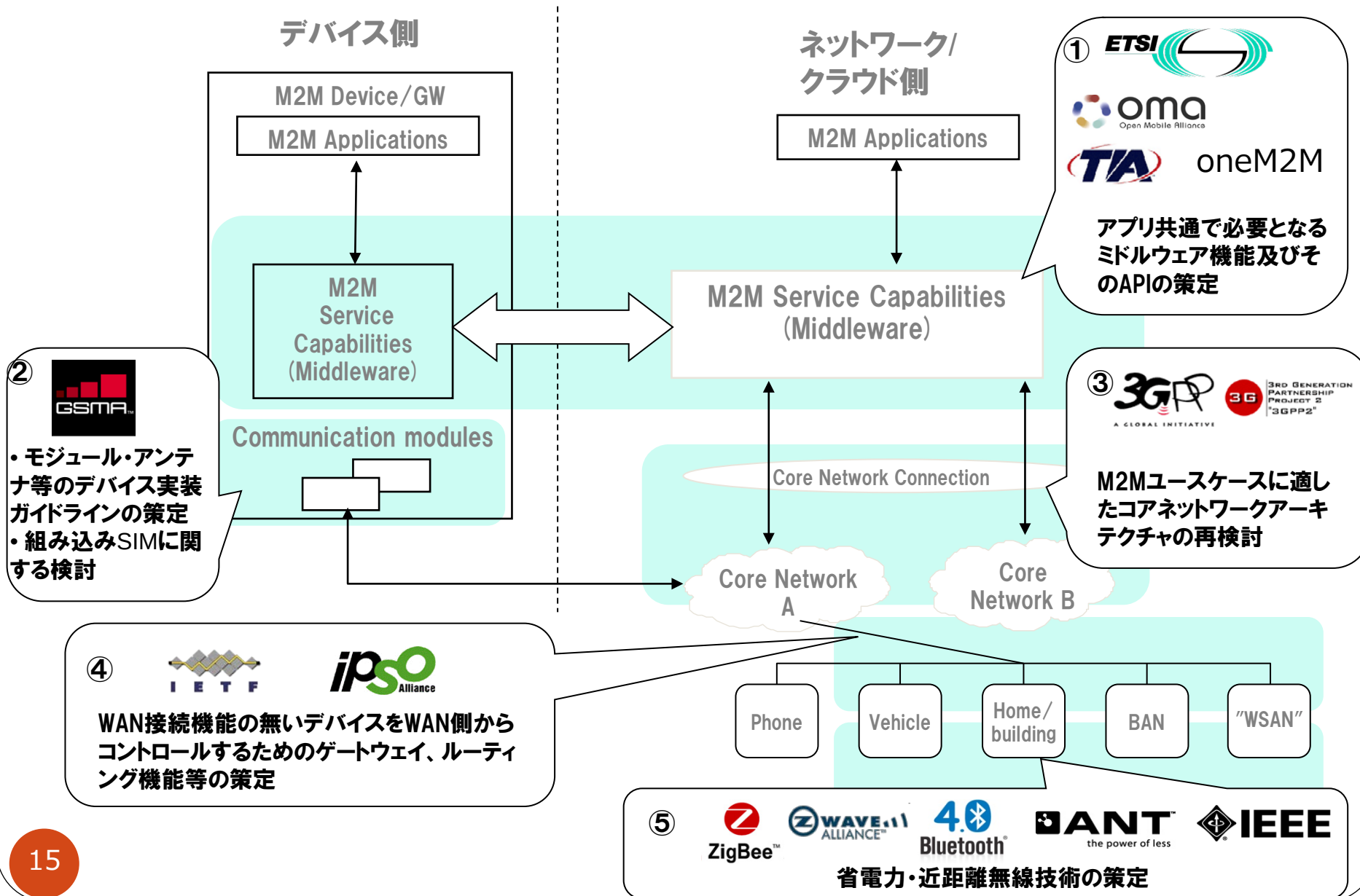


・デバイス種別に特化した情報要素の定義と情報流通方法の定義等。(省電力、メッシュネット)

→各種団体で実施

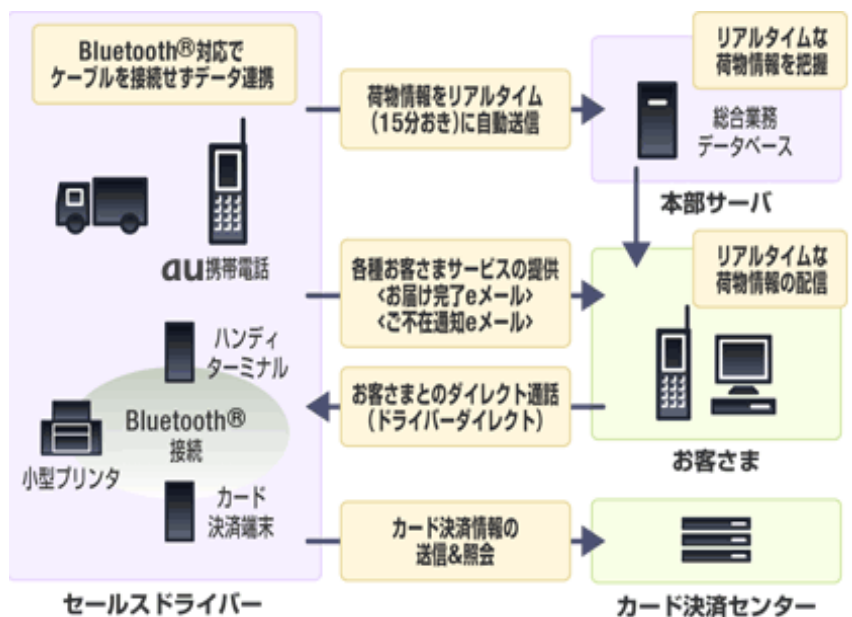
(例: ZigBee Alliance, Z-Wave, UPNP等)

標準化組織の概観

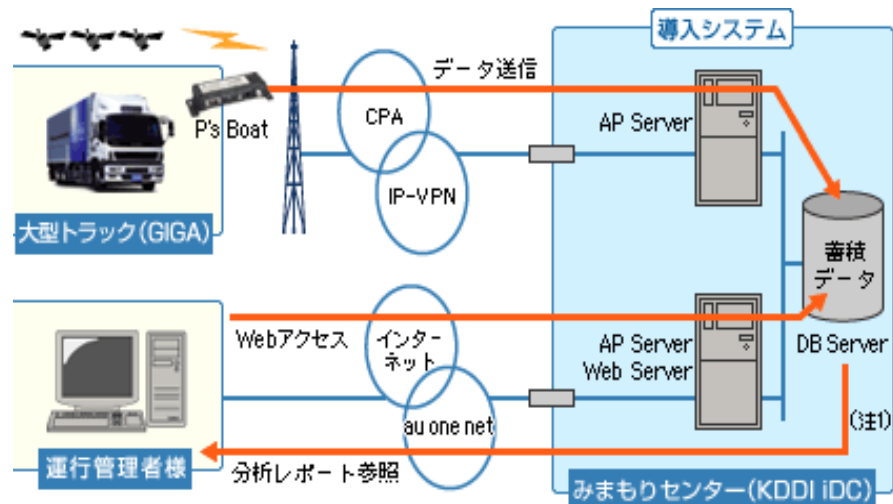


KDDI商用車用テレマティクスサービス お客様事例

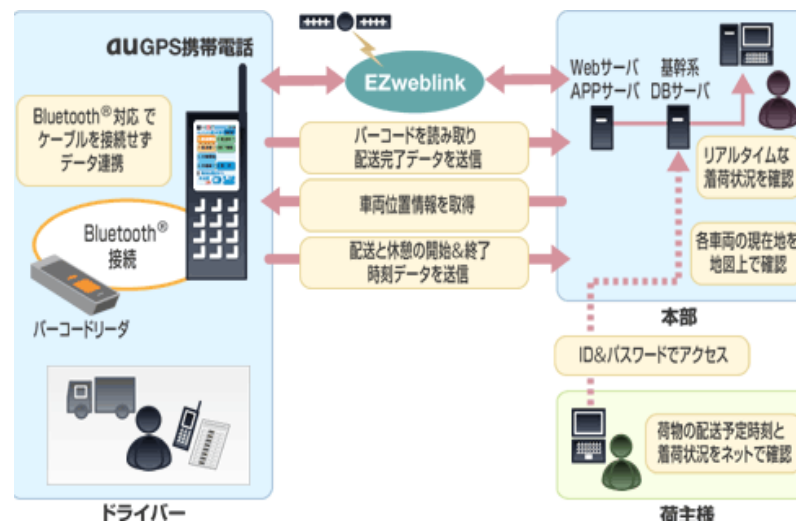
au携帯電話によるデータ通信宅配サービス例



「みまもりくんオンラインサービス」のシステム構成例

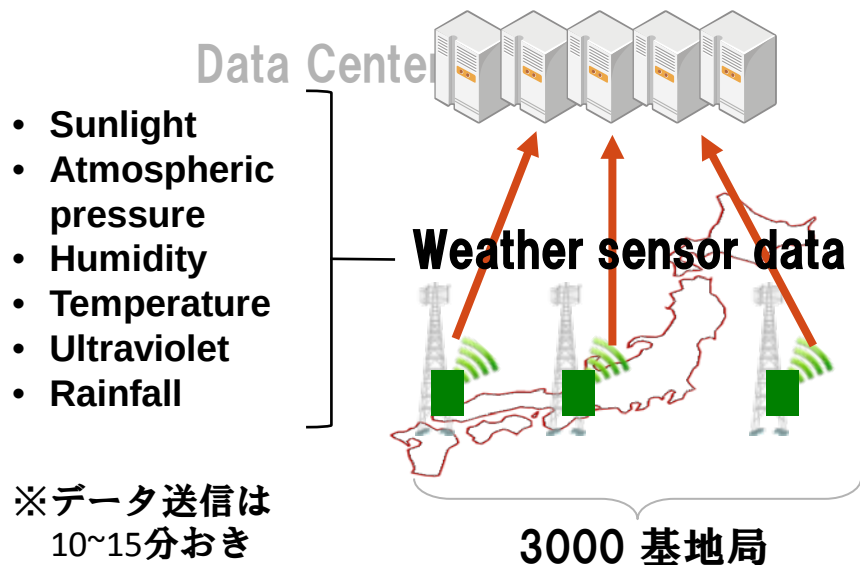


貨物追跡・車両位置管理システム例



気象センサインフラ「ソラテナ」

- 全国3000か所の基地局で1分間隔※で気象センシング
 - 2011/6にウェザーニューズ社とサービス開始（現在はauユーザー向けサービスとして展開中）
- リアルタイム・ピンポイント・高密度な気象情報を提供
- (>>アメダス:全国1300か所、10-60分毎に観測)



現状サービス
(基地局センサが体感気象をつぶやく)

WiMAX X 適用スマートメータ

① WiMAXの電波伝搬試験

擬似メータ

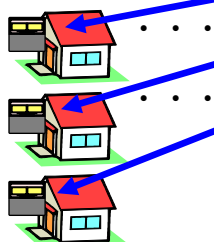


WiMAX

WiMAX基地局



② 複数トラフィックでのデータ伝送試験



検証用
サーバ



③ 実運用をサポートする機能の検証

迂回通信の機能



ZigBee,
特定小電力無線
など

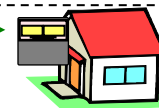


ハンディ
ターミナル



ハンディターミナルとの
直接通信機能

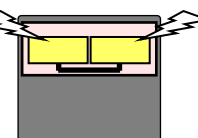
ZigBee,
特定小電力無線など



※③の機能を実現するため、
複数インターフェイスを
備えた擬似メータを製作

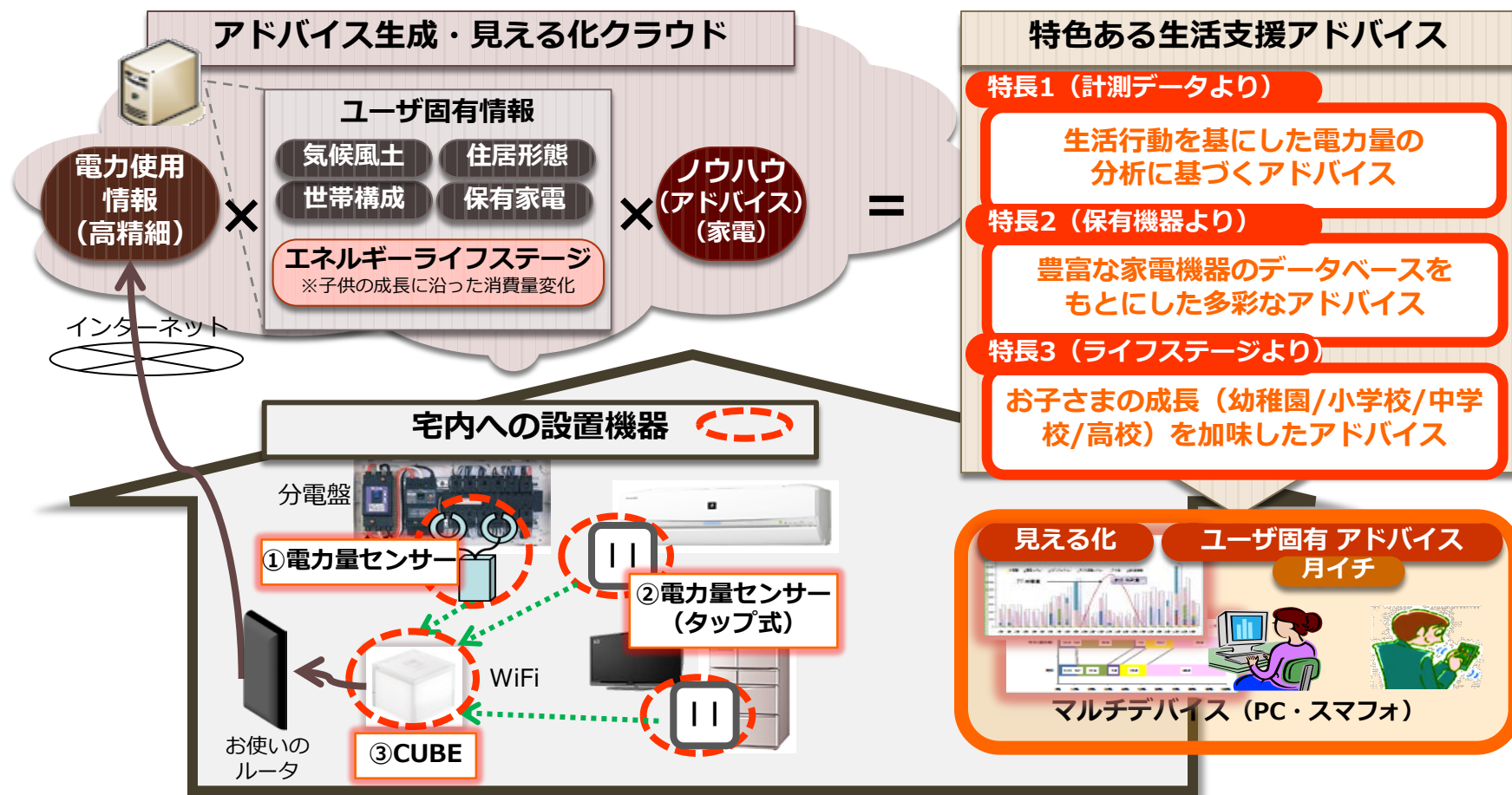
WiMAX

擬似メータ



ZigBee,
特定小電力無線
など

エコビトとらいある



センサデータの分析：移動状態推定

- 技術シーズ：移動状態推定技術
 - ケータイ上のセンサ（加速度、GPS）で7種類の移動状態を推定（走行、歩行、自転車、停止、自動車、バス、電車）
- 日常生活での推定精度・消費電力を実証
 - 被験者20人による1か月間の試験
 - 省電力設計のアプリケーションにより93%の稼働率を達成
 - 推定精度は、歩行83%、自動車81%、電車65%
- 応用
 - ユーザセグメンテーションの高度化
 - 利便性、安心感をもたらす様々なサービスの提供



スポーツのサポート

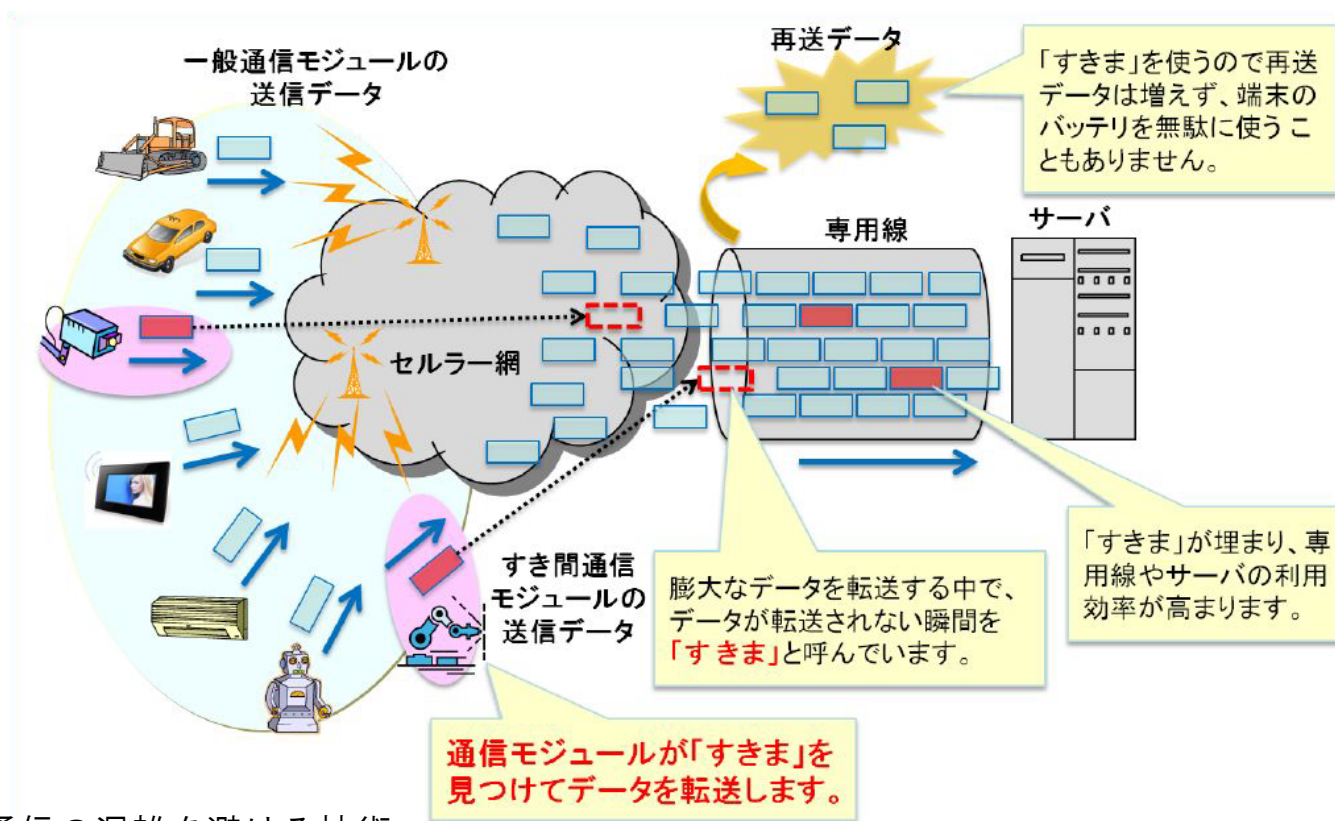


子供がスクールバスに
乗ったら、母親に通知



自動車に乗ったら
助手席ナビを自動的に起動

すき間通信



- M2M通信の混雑を避ける技術
- 統計的手法により、ネットワークの「すきま」が多い時間に自己の通信需要を誘導
- 日・週・月などの単位でまとまった通信をするアプリケーション向けに、スムーズな伝送を確保
- 同時に、ネットワーク負荷を平準化できることを担保
 - 料金施策（相対タリフ）への反映など

センサデバイス管理技術

- スマートフォンをゲートウェイとして、ヘルスセンサやテレマティクス機器等の初期接続設定、プロビ、デバイス管理を実施する技術

1. 目的:

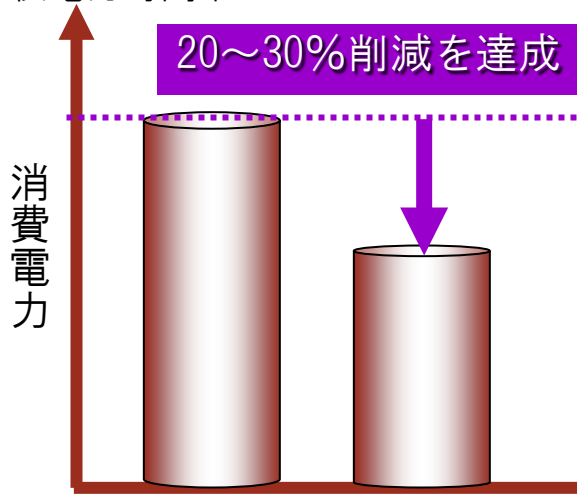
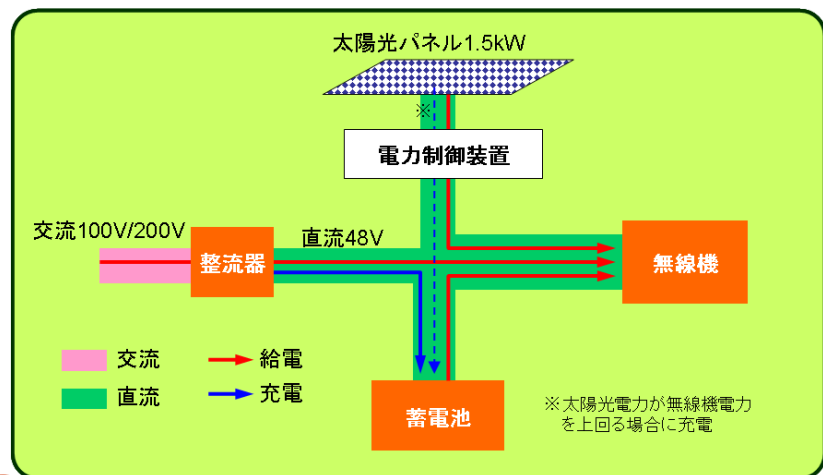
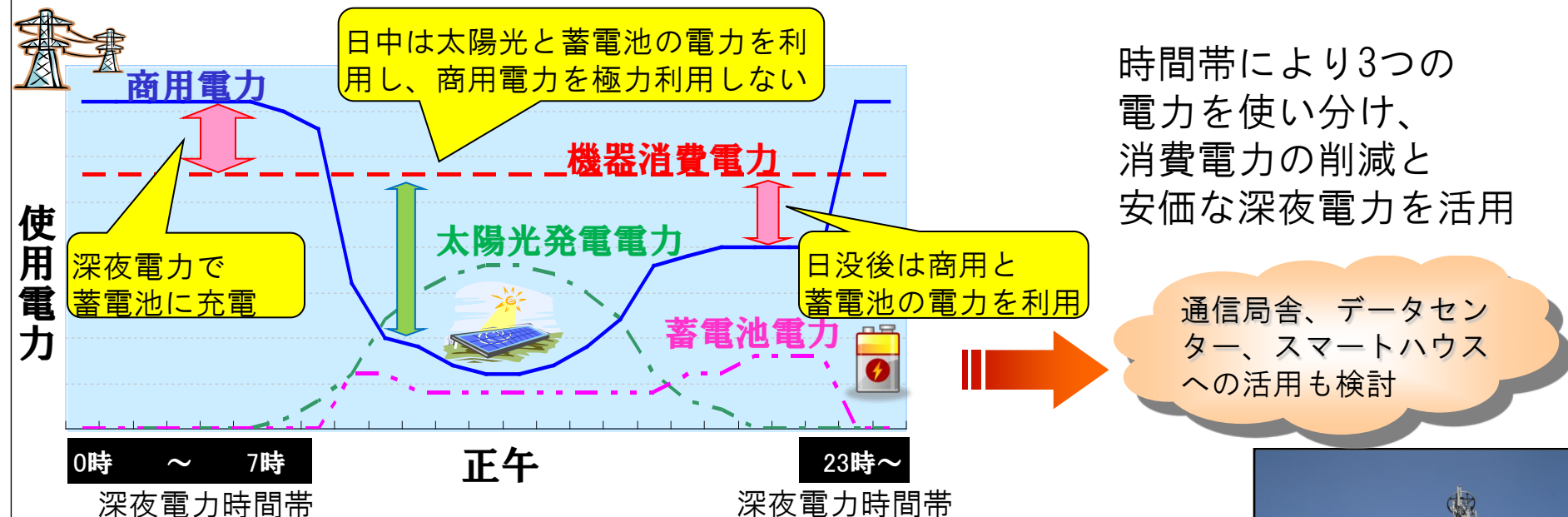
- 車載NWやヘルス機器など、NWや機器の自動設定化が未成熟な分野のゼロアド化を推進する。ビジネスアイデアを容易にビジネス化可能とする基盤を構築し、KDDIの市場を拡大する。

2. 実施内容:

- スマートフォンによるデバイス自動収容機能の開発
 - 対象: テレマティクス・ヘルス分野向けセンサデバイス、ユーザIOデバイス等
 - Plug and Playで自動デバイス発見・初期設定
- スマートフォンによる車載NWの収容手順や、外部アプリ、サーバからのデバイス設定管理APIについて、OMA等で標準化



au基地局向けトライブリッド電力制御方式



スマートフォン-車連携

① 携帯とナビの連携(画面転送)

➤ USB等のケーブル接続からWiFi等でワイヤレス接続へ



② 遠隔携帯操作

➤ 車載機でのパネルタッチにより携帯を遠隔操作

- スマートフォン画面を車載ディスプレイへ出力
- 車載ディスプレイからスマートフォンを遠隔操作
- 既存のスマートフォンのWiFiを利用可
- 車内でスマートフォンをワイヤレス充電

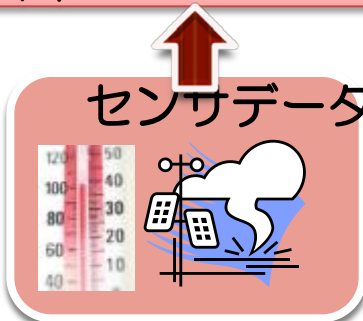
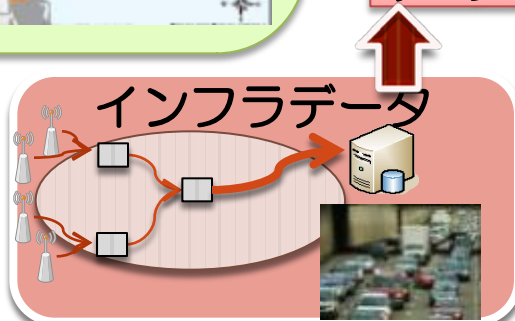
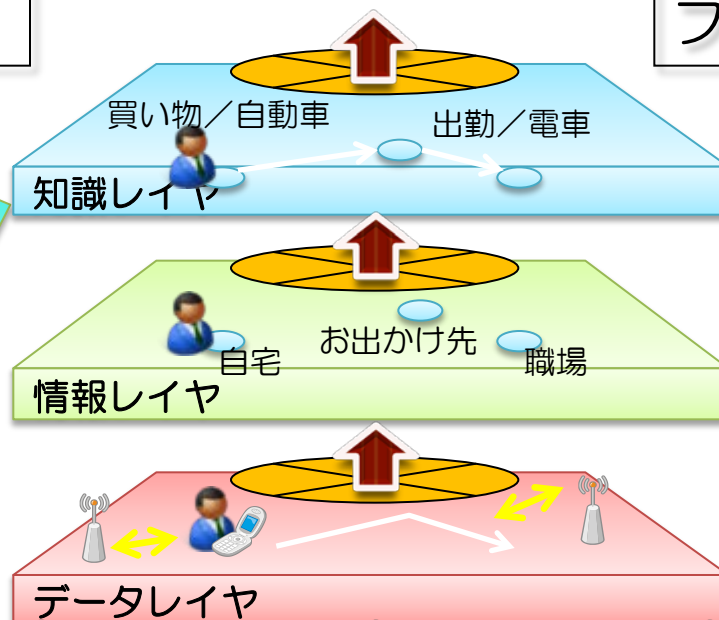
まとめ

ビッグデータがもたらす豊かな社会

個人情報保護しつつ
広く社会で有効活用
するための指針策定

都市計画、防災計画
情報提供サービス

異業種・異分野の
Big Data相互活用
フレームワーク



価値創造・社会発展に向けたBig Data基盤構築