

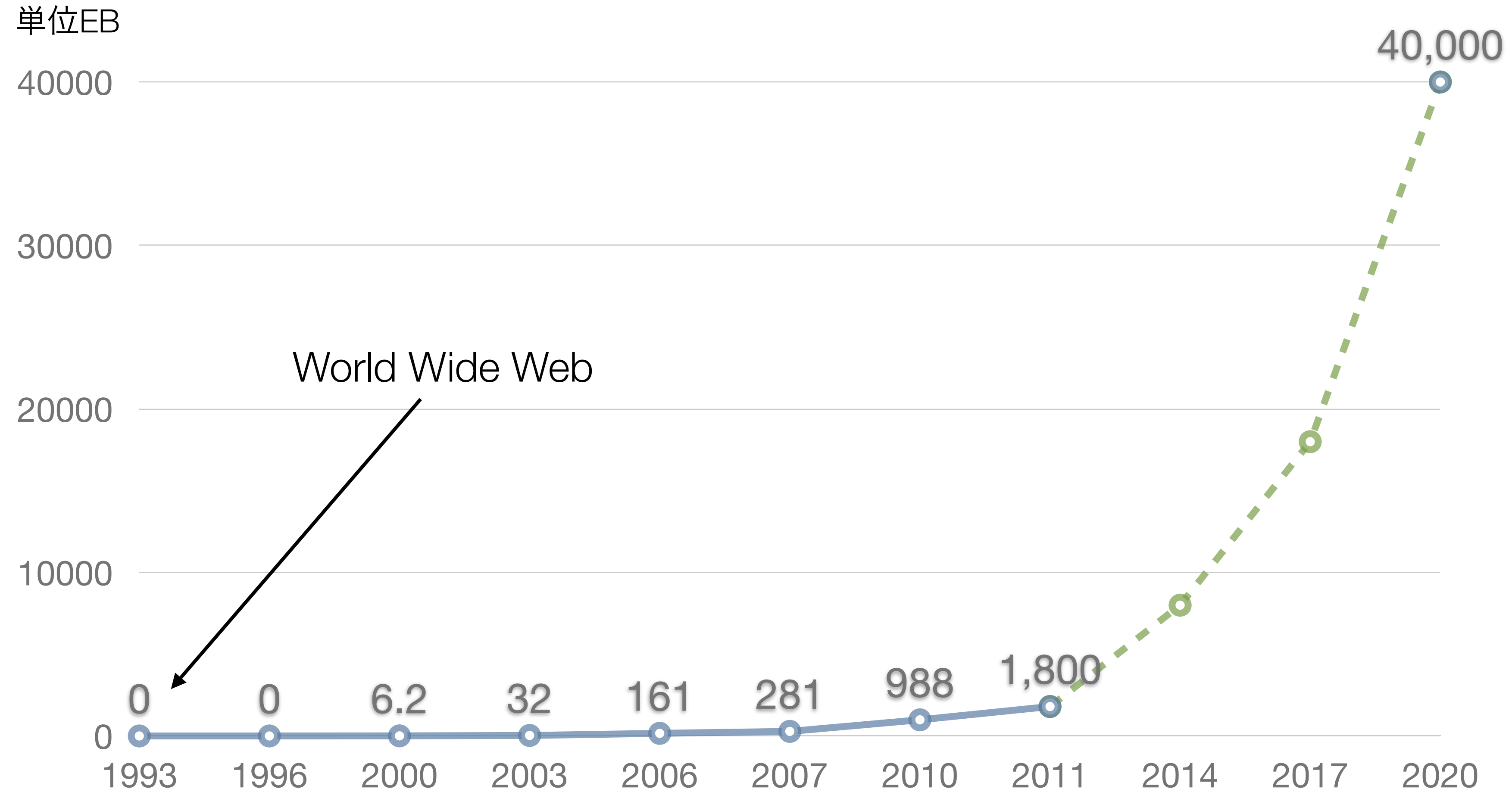
ビッグデータと集合知

情報環境論(8)

ビッグデータ

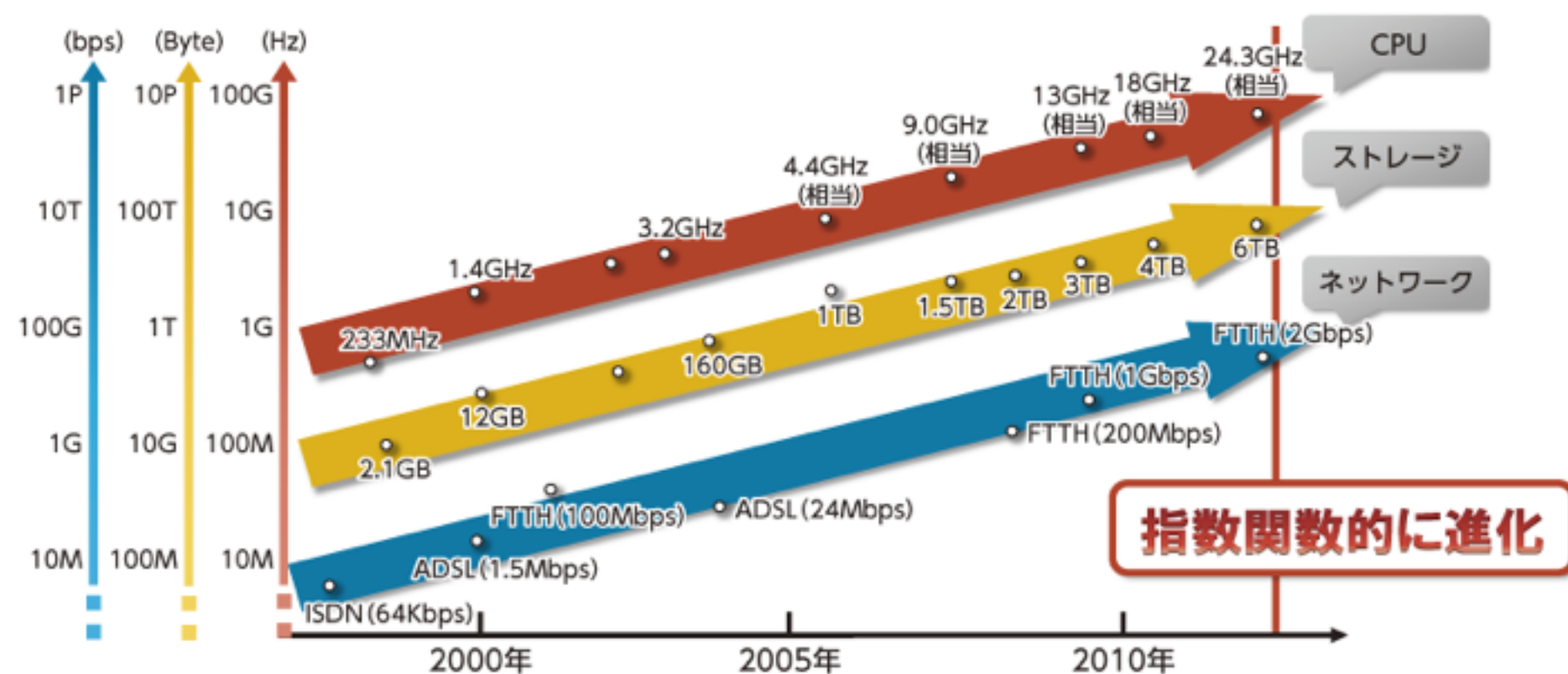
- 量的 → 通常のデータベースソフトウェアで扱えない程の巨大なデータ
 - 数十テラ～数ペタバイト
- 質的 → 多様なデータ
 - 購入履歴やエントリー履歴（オンラインショップやブログなど）
 - マルチメディアデータ
 - ソーシャルメディアデータ
 - センサーデータ（位置・乗車履歴・温度など） → センサクラウド
 - CRM（Customer Relationship Management）システムの会員データなど
- ビッグデータの3V=Volume / Variety / Velocity

世界におけるデジタルデータ流通量の増大

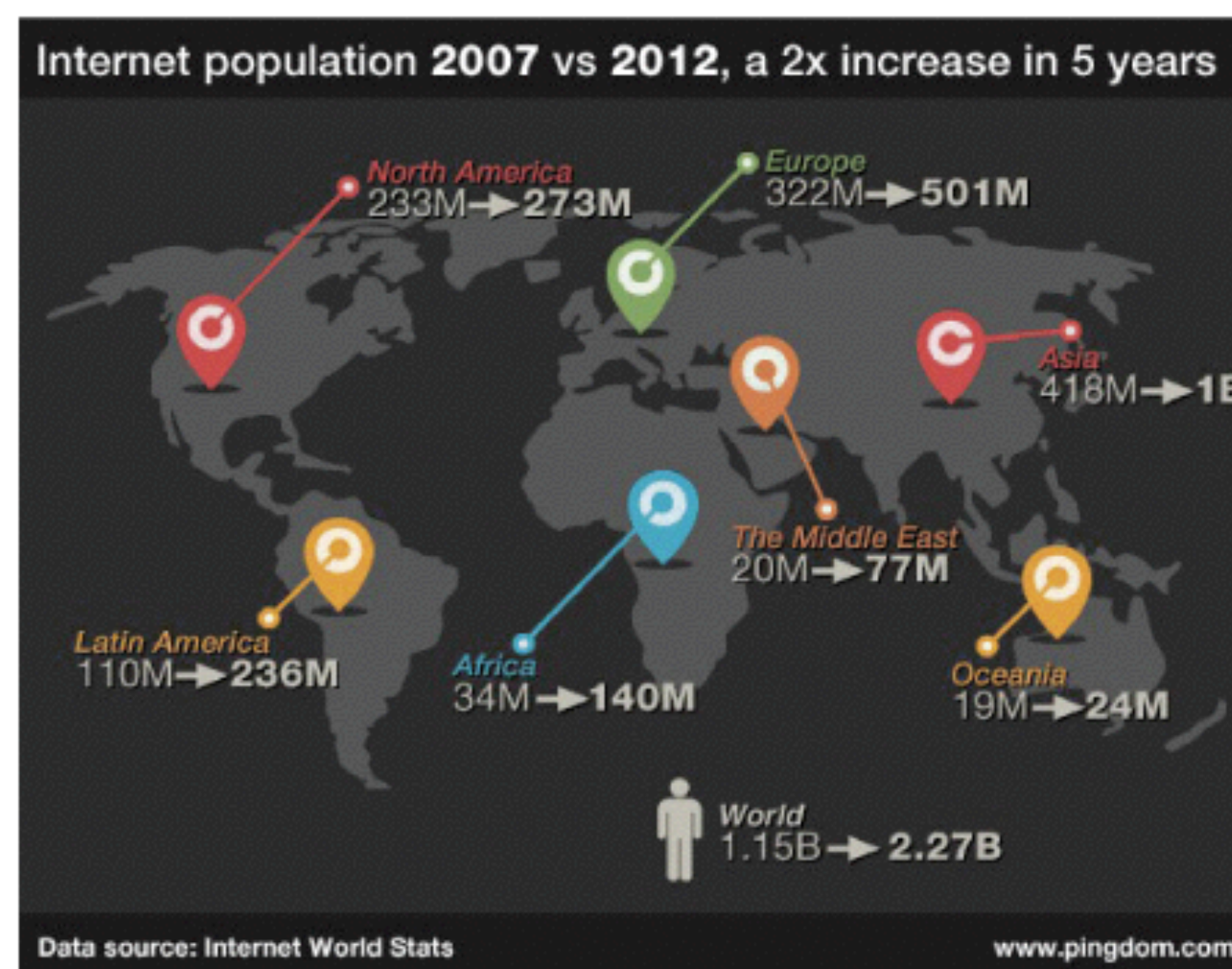


データ流通量増加の背景

1. コンピュータの処理能力の増大



2. インターネットの急速な普及



アフリカ：3400万から1億4000万
- 317%増加

アジア：4億1800万から10億へ
- 143%増加

欧州：3億2200万から5億100万へ
- 56%増加

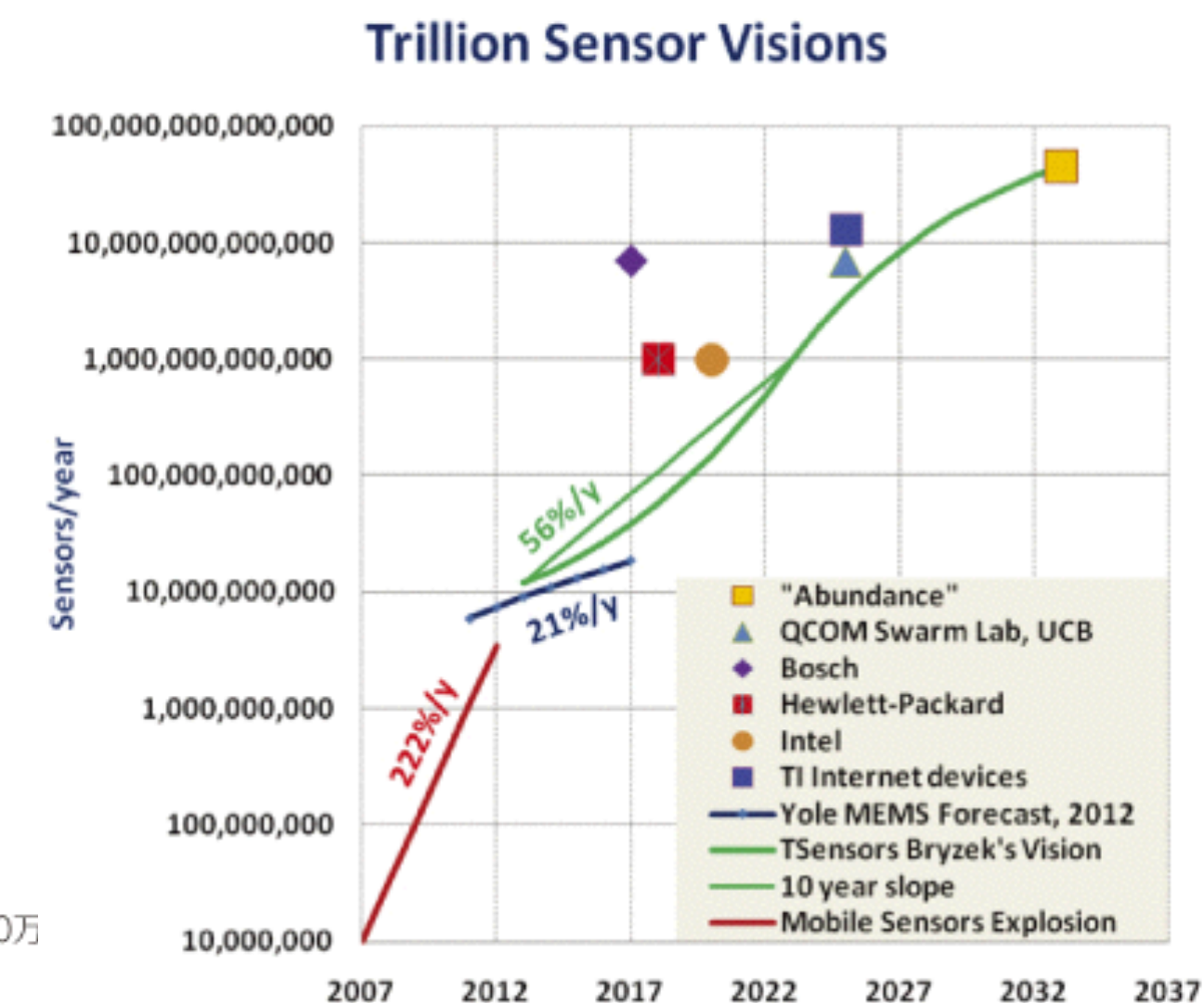
中東：2000万から7700万へ
- 294%増加

北米：2億3300万から2億7300万へ
- 17%増加

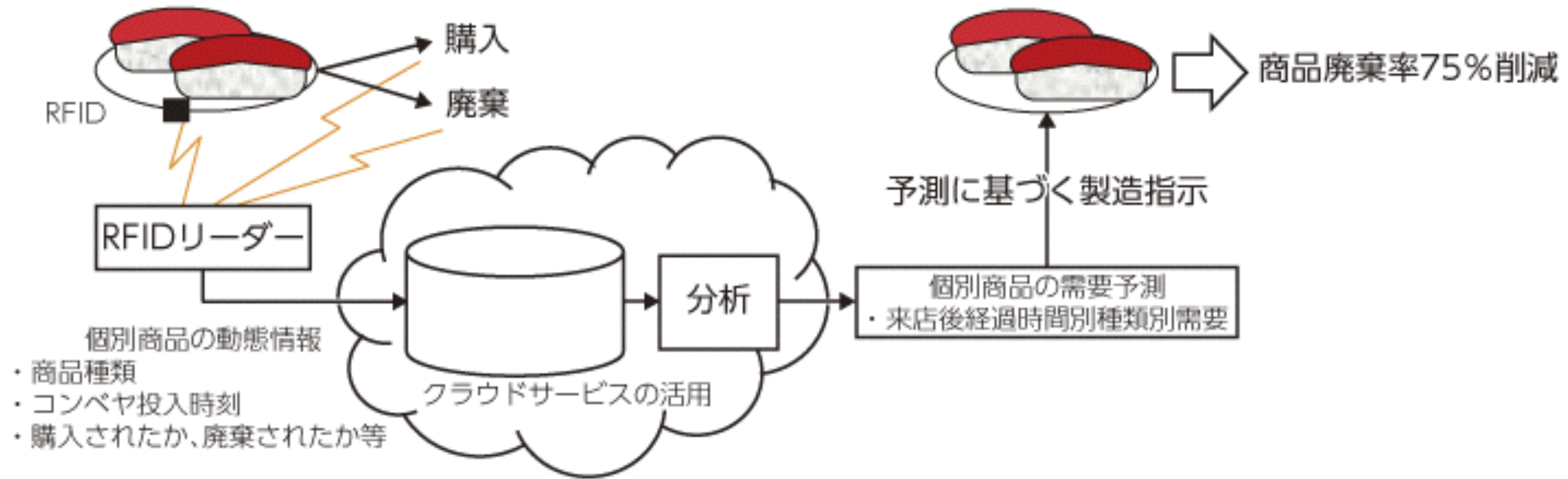
中南米：1億1000万から2億3600万へ
- 114%増加

オセアニア：1900万から2400万へ
- 27%増加

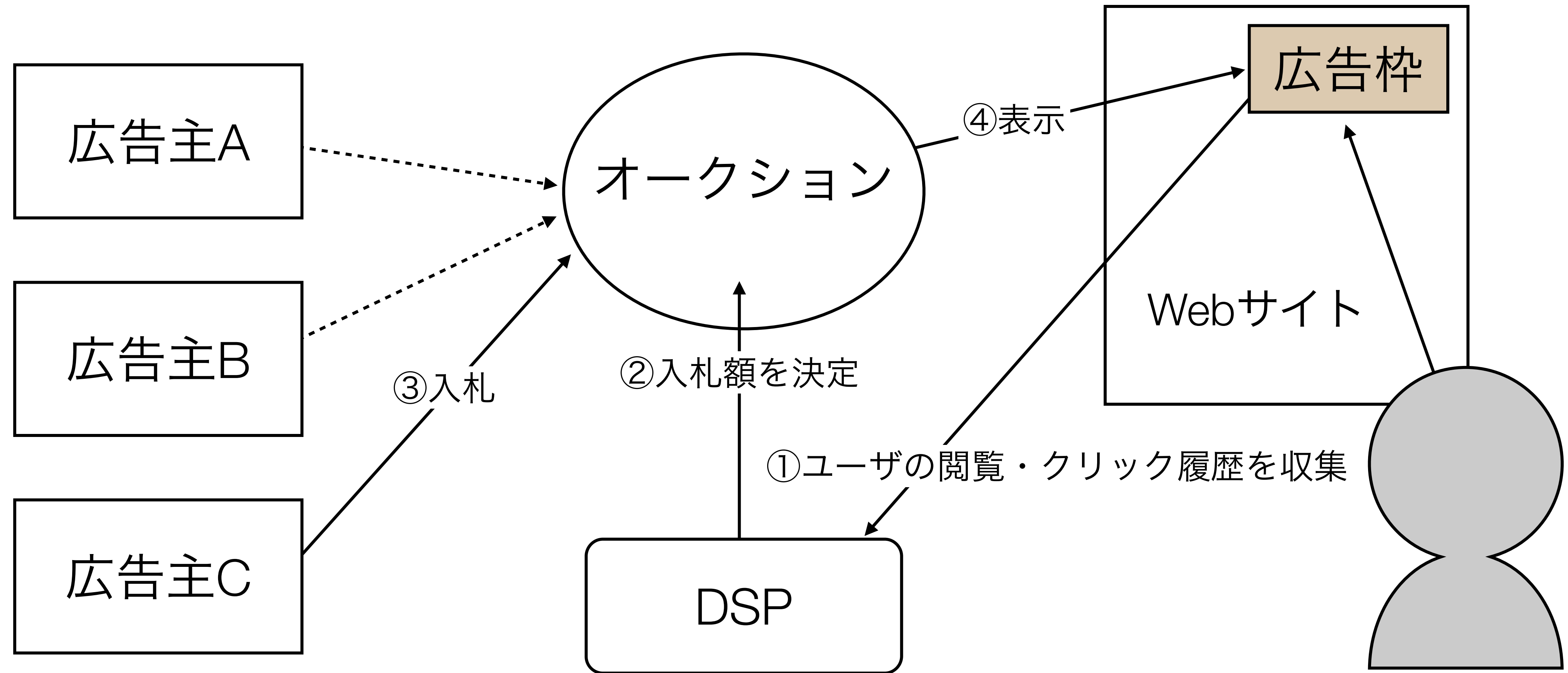
3. センサーの普及



ビッグデータの活用(1) スシロー



ビッグデータの活用(2) インターネット広告



インターネット広告

- SSP : Supply-Side Platform
 - ✓ 広告枠に提示する広告を選択する (広告枠側)
 - ✓ 広告がクリックされると広告業者から広告代が支払われる
- DSP : Demand-Side Platform
 - ✓ 広告が提示される広告枠を選択する (広告主側)
 - ✓ 広告主は広告枠を買う
 - ✓ 広告がクリックされると広告主は広告業者に広告代を支払う

集合知

- Collective Intelligence \Leftrightarrow Wisdom of Crowds
- 元々は生物学（昆虫学）の用語
 - 比較的単純な（知性を持たない）個体の集団が知性を持つように見える
ex) 蟻の集団, 粘菌コロニー
- 社会学, 経済学, コンピュータ科学への拡張
 - 多くの人々の知識や情報を集めて新しい知見を得る
 - ≡ データマイニング

データマイニング

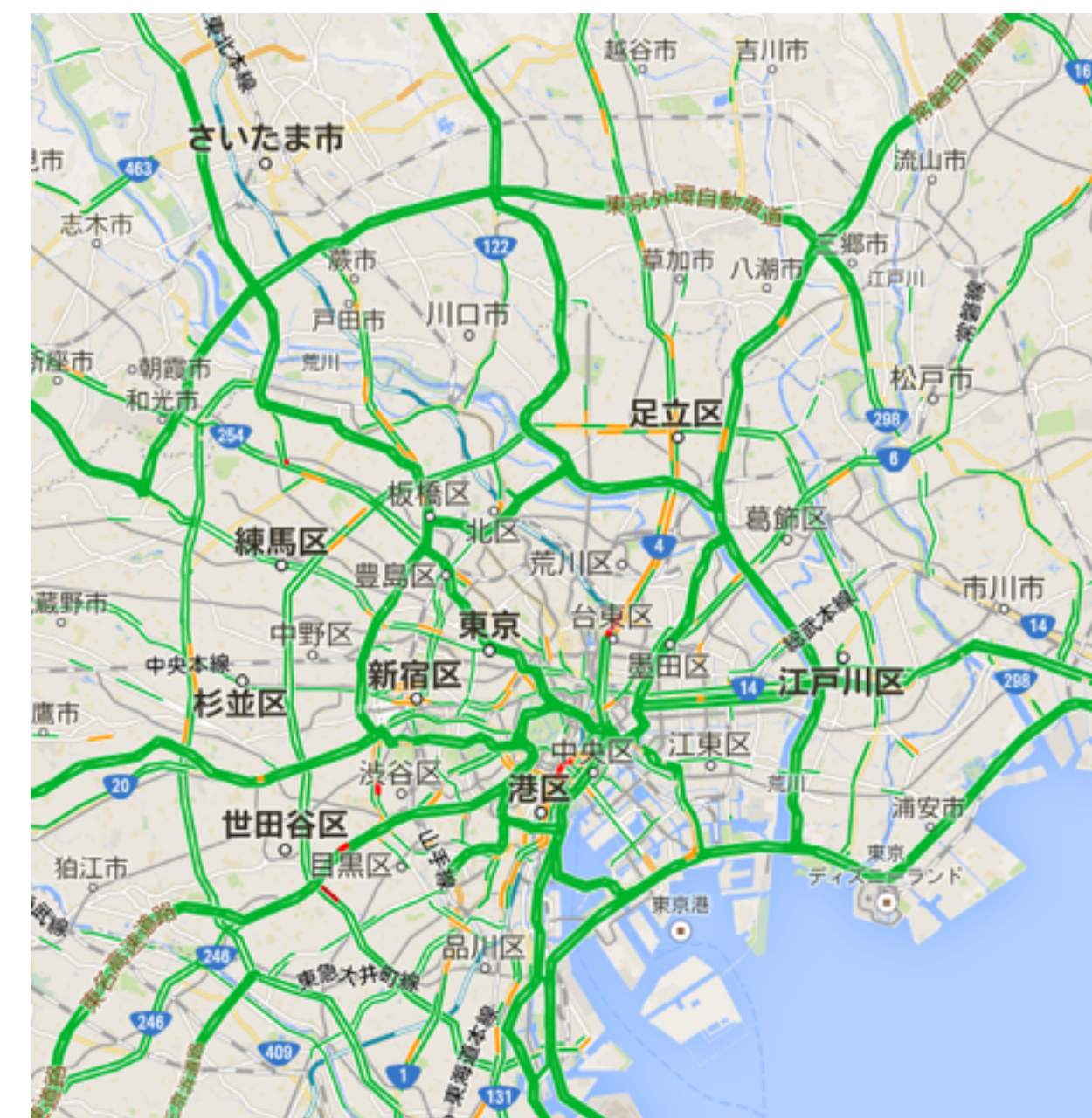
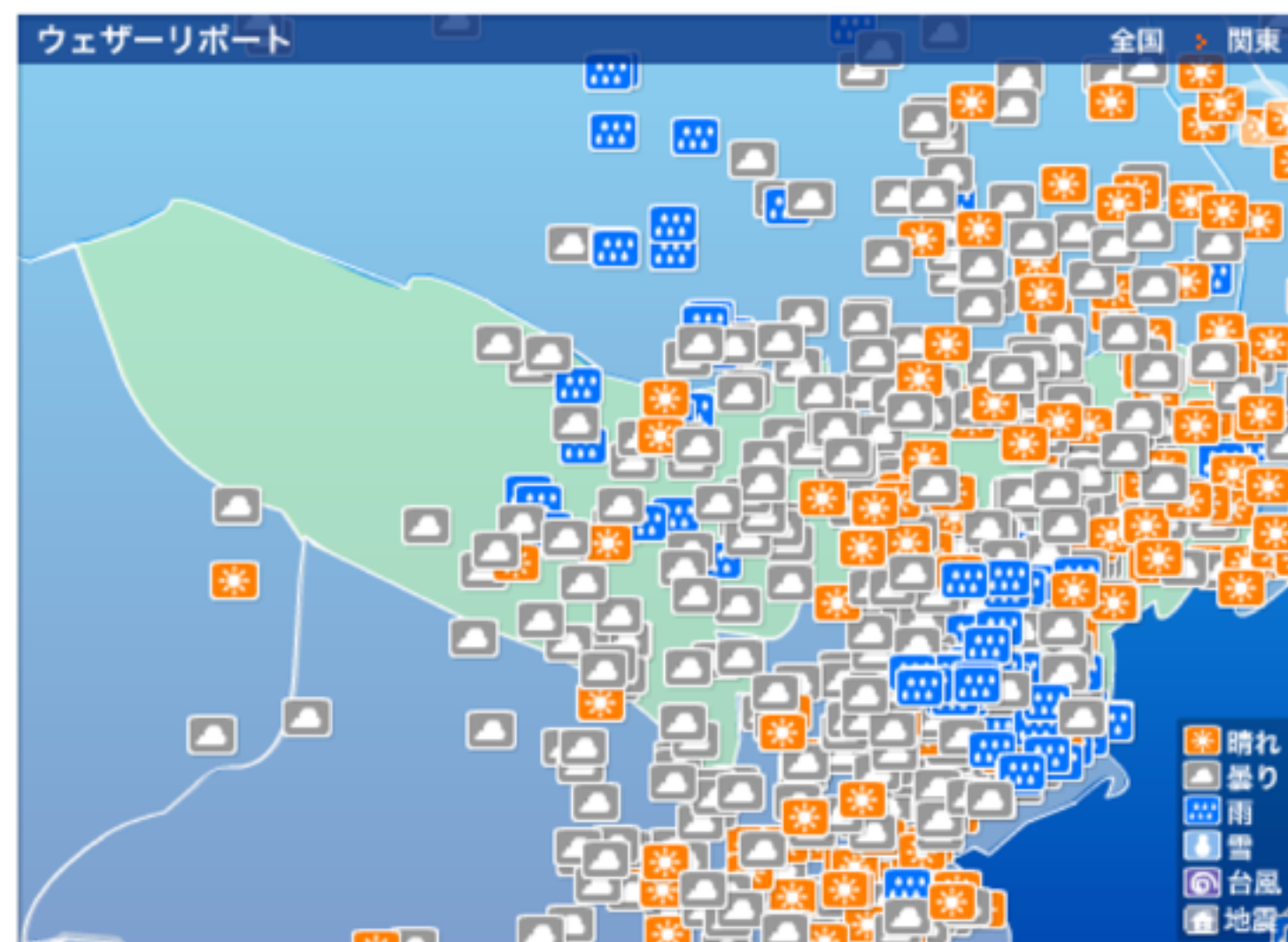
- 大量のデータから知識を取り出す技術・手法の包括的な呼称
 - ✓ 特にテキストデータが対象 → テキストマイニング
 - ✓ 特にWebページが対象 → ウェブマイニング
- 統計学・人工知能・機械学習・パターン認識などの手法を利用

ビッグデータと集合知

- Wikipedia
 - ✓ 集合知によるフリーなオンライン百科事典のプロジェクト
 - ✓ 延べ90万人（日本語版） / 2300万人（英語版）が参加
- SETI@home
 - ✓ 家庭用PCの余剰計算能力を利用したSETIの補助プロジェクト
 - ✓ 世界300万台以上のPCが参加
- キュレーションサービス
 - ✓ Togetter, NAVERまとめ, Redditなど
- ナレッジベース
 - ✓ Yahoo!知恵袋など
- 協調フィルタ
 - ✓ Amazonの書籍推薦

ソーシャルセンサ

- ビッグデータなどを分析してイベントの検出に利用
 - ✓ 渋滞予測 (Google Mapなど)
 - ✓ 電車の遅れの把握
 - ✓ 視聴率 (Twitter TVエコー)
 - ✓ 天気概況
 - ✓ 地震の被害推定
 - ✓ 株価変動
 - ✓ 事件の予測・防止
 - ⇔ スノーデン事件



スノーデン事件

- エドワード・スノーデン（元CIA・NSA局員）
- NSAが行ったインターネット・電話などの傍受を暴露（2012～2013年）
- バンドレス・インフォーマント（NSAの監視ツール）
 - ✓ クラッキングではなくAPIベースの監視ツール
 - 米国のIT企業が協力

