

ビッグデータ

ビッグデータとは

- 単に「大量のデータ」ではなく次のVのいずれか、あるいは複数が極めて高いもの
 - Volume: データの量
 - Variety: データの種類（非構造化・非定型データ）
 - Velocity: データの頻度（日々蓄積されるリアルタイムデータ）

- ビッグデータの例
 - 銀行の取引履歴 ➡ 構造化・定型データ
 - 銀行取引が行われたときの利用者の環境・境遇 ➡ 非構造化・非定型データ
 - ライフログ ➡ 非構造化・非定型データ，リアルタイム
 - 各種センサからの情報 ➡ 非構造化・非定型データ，リアルタイム

データ蓄積・分析の歴史

- 1970年代 電算化
 - ホスト勘定系システム, IBMメインフレーム
- 1980年代 コモディティ化
 - MS-DOS, スプレッドシートソフトウェア
- 1990年代 オープン化
 - リレーショナルデータベース, インターネット
 - データウェアハウス：RDBによる顧客情報の収集（蓄積するまえに整理）

- 2000年代 データマイニング
 - Google
 - とりあえずデータはためるもの。欲しいものは探す・発掘するもの
- 2010年代 ビッグデータ
 - スマートフォン, IoT デバイス, AI
 - データでもどこでもあるもの
 - データは分析しないもの

ビッグデータは何を生むのか

- なぜこれまで捨てられてきた情報に価値がある？ ➡ 捨てても困らない
- 「原因と結果」の因果関係がわかるか？ ➡ 因果関係はおそらくわからない
- ビッグデータを**分析**しても**何も生まない**

- 「データマイニング」は価値のある（金となる）**因果関係を分析したい**
 - マイニングで発掘された情報そのものは価値があるが生き方そのものは変わらない
 - 「おむつとビールの売り上げに相関がある」
- 「ビッグデータ」は価値のある**モデルを構築したい**
 - 生き方そのものの変化
 - 「必要なものが適切なタイミング自動的に注文されて送られてくる」