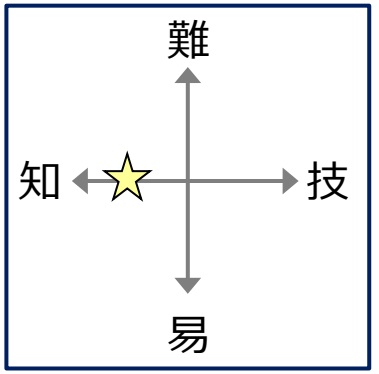
総務省　ICTスキル総合習得教材



［コース1］データ収集

1-1：IoTとデータ利活用の全体像

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_1_1.pdf

	1	2	3	4	5
［コース1］データ収集	●				
［コース2］データ蓄積					
［コース3］データ分析					
［コース4］データ利活用					

本講座の学習内容［1-1：IoTとデータ利活用の全体像］

【講座概要】

- IoT（Internet of things）の基本概念に加えて、IoT機器の事例を紹介します。
- ICT（Information and Communication Technology：情報通信技術）を用いたデータの利活用の全体像と4つの段階の概要を説明します。
- 世界におけるIoTの普及に関する予測と各国のICTの取り組みを示します。
- 日本におけるIoTに関する取り組みと産業における利活用例を示します。

【講座構成】

座学

- [1] IoTとその事例
- [2] ICT・データ利活用の全体像
- [3] 世界におけるIoTの普及と各国の取り組み
- [4] 日本のIoTに関する取り組みと産業での利活用

【学習のゴール】

- ✓ IoTの基本概念と機器の事例を把握する。
- ✓ データ利活用の4段階の概要を理解する。
- ✓ 世界的なIoTの普及動向と各国におけるICT活用に関する取り組みを把握する。
- ✓ 日本のIoTに関する取り組みと産業における利活用例を紹介できる。

IoT（Internet of Things）

◆IoTは、様々なモノがインターネットに接続することを指しています。

- **IoT（アイオーティー）**は、**Internet of Things（インターネット オブ シングス）**の略で、「様々な物がインターネットにつながること」「インターネットにつながる様々な物」を指しています。
  - インターネット(internet）は、inter(相互の)とnet(ネットワーク)からできた言葉で、ネットワーク同士をつなぐ世界全体の情報通信網です。
- IoTは、日本語で**「モノのインターネット」**と訳され、PCに限らず様々なモノがインターネットにつながります。
  - PCでのインターネットの利用方法は「ウェブサイトの閲覧」「メールの送受信」が中心ですが、IoTにおけるインターネットの利用方法は、より広範です。
- IoTが普及すれば、衣服、白物家電、家屋といった様々な物がインターネットにつながります。
  - インターネットにつながる「衣服」は身につけるIoTの「ウェアラブルデバイス」の一種です。「ウェアラブルデバイス」に関しては、講座1-2にて説明します。

インターネットにつながるものとして
イメージされやすい物

IoTの普及

今後一層インターネットにつながる物のイメージ

□ この教材のタイトルでもある**ICT（アイシーティー）**は、**Information and Communication Technology（インフォメーション アンド コミュニケーション テクノロジー）**の略で**情報通信技術**を指しています。

家庭用IoT機器の事例

◆IoTは情報を送信するタイプと受信するタイプがあり、家電製品としても発売されています。

- インターネットとつながったIoT機器には、**情報を発信するタイプ**もあれば、**情報を受信するタイプ**もあります。
- 象印マホービン株式会社が発売しているIoT電気ポットの「iポット」は電源を入れたり、給湯したりすると、**情報発信型IoT**として、離れて暮らしている家族のスマートフォン等へ動作状況が伝達されます。
  - IoT機器からの動作状況等の通知によって、お年寄りや子供の状態を把握できるサービスは「見守りサービス（安否確認サービス）」と言われます。
- パナソニック株式会社が発売しているロボット掃除機の「ルーロ（MC-RS800）」は、**情報受信型IoT**として、外出先からでもスマートフォンで動作させることができます。
  - ルーロには、センサーや人工知能が搭載されており、ロボットとして考えることができます。ロボットに関しては、講座1-4にて説明します。

iポット（情報発信型IoT）

【出所】iポット［象印マホービン株式会社］
http://www.mimamori.net/product/

ルーロ（情報受信型IoT）

【出所】ルーロ（MC-RS800）［パナソニック株式会社］
https://panasonic.jp/soji/rulo.html

IoTによる扉のスマートロック

◆建物のIoT化の事例として、扉の鍵を管理・制御するスマートロックが挙げられます。

- 建物のIoT化の事例として、扉の鍵の施錠・解錠をスマートフォンから管理できるスマートロックが挙げられます。
  - 情報通信技術を活用して効率的な利用ができる製品やサービスには、「スマートフォン」のように名詞の前に「スマート」という言葉をつけることがあります。
- Qrio株式会社が販売している「Qrio Smart Lock（キュリオ スマートロック）」を扉の内側に取り付けることで、スマートフォン等から遠隔で鍵の状況を確認できるとともに、施錠・解錠の操作ができます。
- スマートロックを使うことで利用者・時間を限定した部屋の利用権の設定・管理が容易になるなど、IoTは**シェアリングエコノミー**の活性化にも貢献します。
  - 平成29年版 情報通信白書［総務省］において、「シェアリング・エコノミーとは、個人等が保有する活用可能な資産等を、インターネット上のマッチングプラットフォームを介して他の個人等も利用可能とする経済活性化活動である」と説明しています。

【出所】平成29年版情報通信白書［総務省］
http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/html/nc112220.html

Qrio Smart Lockの鍵への取り付けとスマートフォンによる操作

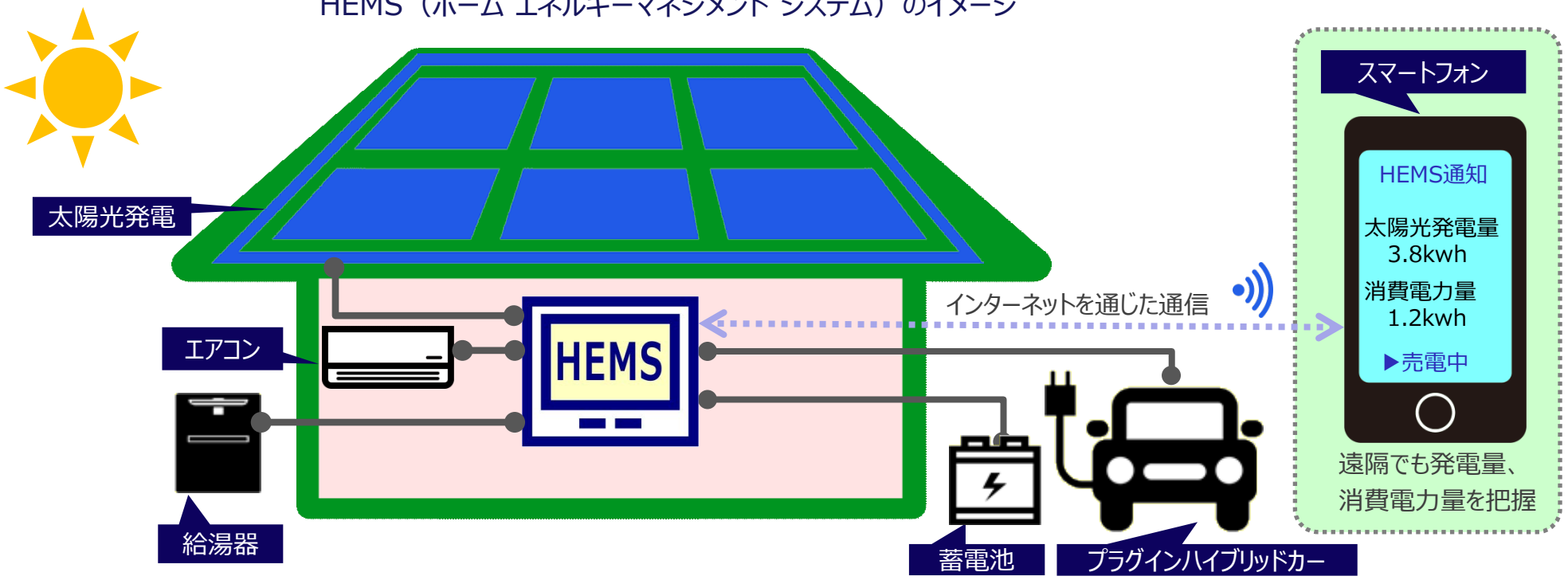
【出所】Qrio Smart Lock［Qrio株式会社］https://qrio.me/smartlock/q-sl1/hardware/

IoTによる家屋の電力制御

◆建物のIoT化の事例として、家屋の電力制御のHEMSが挙げられます。

- 家屋の電力制御システムである**HEMS（ヘムス）**やオフィスの電力制御システムの**BEMS（ベムス）**は、建物における発電および電力利用と電力市場に関する情報に基づいて、電力使用量や売電量を制御します。
  - HEMS（ヘムス）は、Home Energy Management Systemの略、BEMS（ベムス）は、Building Energy Management Systemの略です。
- HEMSやBEMSは、インターネットを通じて電力価格を受信し、電力価格が安い時に蓄電池へ電気を貯める一方で、電力価格が高い時には家庭用蓄電池から売電するなど、効率的な電力利用に貢献します。
  - HEMSでは、屋根にとり付けた太陽光発電の発電量や電力消費量をインターネットを通じてスマートフォンなどで把握することもできます。

HEMS（ホーム エネルギーマネジメント システム）のイメージ

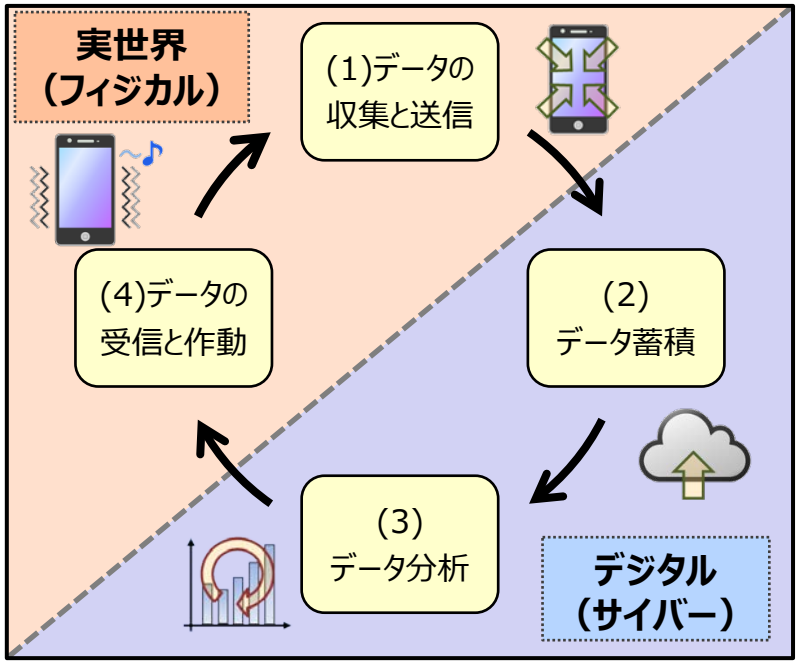


ICT・データの利活用の4段階

◆このパートでは、ICT・データの利活用に関する4段階の概要を説明します。

- IoT機器は「データの収集と送信」や「受信データに基づく作動」を行いますが、ICT（情報通信技術）・データの利活用において、この工程の間には「データの蓄積」「データの分析」が入り、4段階で構成されることがあります。
  - IoTは実世界においてデータ収集や作動を行いますが、データの蓄積やデータの分析はコンピューターの中のデジタルの世界で行われることが一般的です。
  - IoTにとり付けたセンサー等により実世界（フィジカル）のデータを収集・観測し、デジタル（サイバー）空間にてデータの処理・分析を行い、データから得られた価値を実世界に還元することは、CPS（Cyber-Physical System：サイバーフィジカルシステム）とも呼ばれます。
- **（1）データの収集、（2）データの蓄積、（3）データ分析、（4）分析・データに基づく作動**のそれぞれの段階においては、注目されている製品・技術・サービスがあります。
  - データの蓄積においてはクラウド、データの分析においてはAI（Artificial Intelligence：人工知能）や機械学習が注目されています。

データの利活用の4段階



利活用の各段階において注目されている製品・技術・サービス

ICT・データ利活用の段階	注目されている 製品・技術・サービス
（1）データの収集	IoT機器 ［センサー］
（2）データの蓄積	クラウド ［仮想化・分散技術］
（3）データの分析	AI（人工知能）・機械学習
（4）分析・データに基づく作動	IoT機器 ［アクチュエーター・ロボット］

1段階目：センサー等によるデータの収集

◆IoT機器は、取り付けられたセンサー等で様々なデータを収集することができます。

- IoT機器には**センサー**をとり付けることができ、人間の五感（視覚、聴覚、味覚、嗅覚、触覚）で感知できるような**データの収集**ができます。
  - カメラや照度センサーをとり付ければ視覚情報のデータ収集を行い、マイクや音量センサーをとり付ければ聴覚情報のデータ収集ができます。
- 人間がキーボードなどから入力した情報、機械の稼働状態を記録しているログ情報も収集するデータになり得ます。
  - PCへの入力に利用されるキーボードやマウスは、圧力（触覚情報）に反応するセンサーとして考えることができます。

- IoTに取り付けるセンサーでは、人間の五感では感知できない電波や紫外線なども感知、計測することができます。
- 株式会社ウェザーニューズが販売し、情報収集に利用している簡易気象観測器WxBeacon2（ウェザービーコン2）は、気温・湿度・気圧・明るさ・紫外線・騒音の6種のデータを収集し、スマートフォンでの表示が可能です。

6種類のセンサーデータを収集するWxBeacon2とスマートフォンにおけるデータ表示

※WxBeacon2 のデータは、およそ1分毎に読み込まれます。

気温 26.7℃	湿度 55.3%
照度 79lx	紫外線指数 0.0
気圧 1003.4hPa	騒音 38.2dB
不快指数 74.7	熱中症危険度 23.4℃

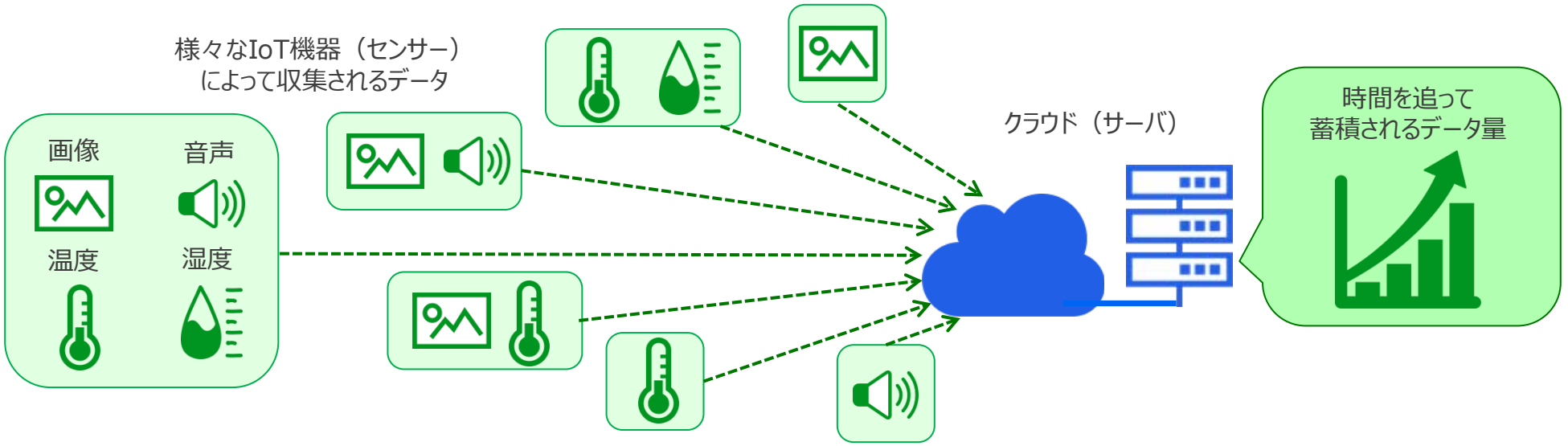
【出所】WxBeacon2［株式会社ウェザーニューズ］
https://weathernews.jp/smart/wxbeacon2/

2段階目：クラウド等へのデータの蓄積

◆データはクラウド等に蓄積されることで、より広範で大規模な活用が可能になります。

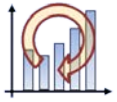
- IoT機器が収集したデータは、**クラウド**等の**インターネット上のサーバ**に送信されて保存されることで、円滑なデータの集約と蓄積が可能になります。
  - サーバは要求に応じてサービスを提供するコンピュータを指し、ここでは要求に応じてデータを保存するコンピュータを意味しています。
  - クラウドは講座2-5において説明するようにデータ蓄積以外にも活用されていますが、ここではデータを蓄積するサーバの一種とのお考えください。
- センサー等を搭載した複数のIoT機器で得られた情報を集約することで、様々な種類の情報がサーバに集まります。
  - IoT機器によって収集する情報は、画像データ、音声データ、温度や湿度などの数値データなど様々です。
- クラウド等のサーバでは、大容量のデータの保存が可能であり、長期間の**膨大なデータを蓄積**できます。
  - IoT機器は記憶するパーツを持たないか、持っていても記憶容量が小さいため、長期的なデータ蓄積にはクラウド等のサーバを活用することが一般的です。

IoT機器（センサー）からクラウド（サーバ）へのデータの集約・蓄積のイメージ



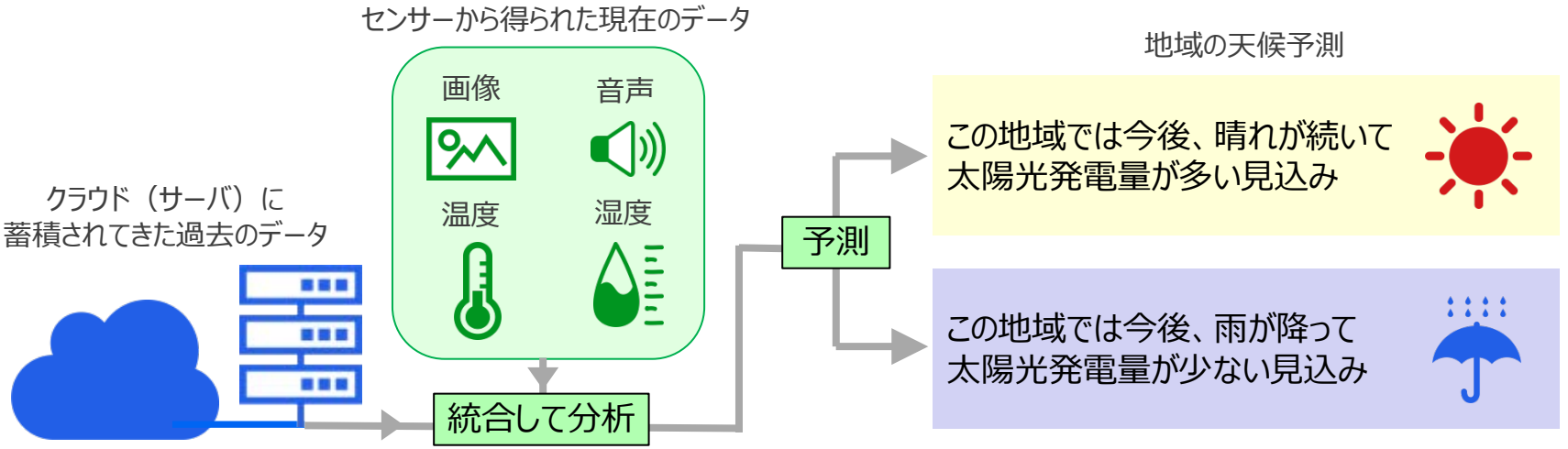
3段階目：人間や人工知能によるデータ分析

◆蓄積されたデータは、分析を行うことによって、その後の有効活用につながります。



- 蓄積された膨大な**データを分析**することで将来予測、課題解決のヒントが得られることがあります。
  - IoT機器等で収集されたデータは、多様、大規模、高頻度（リアルタイム性）といった特性を持つケースが多く、講座3-1で示すようにビッグデータと呼ばれることがあります。ビッグデータの特性やデータ分析については、コース3において説明します。
- 蓄積されたデータと現在のセンサーデータを統合して分析することで、即座に機器の作動、人間の行動に反映するべき近い将来の予測をすることも可能になります。
- 分析作業は従来、人間が行うことが一般的でしたが、**AI（人工知能）**などによって、コンピュータが自動的に行うケースも増えてきています。
  - 人工知能や人工知能に関する分析技術である機械学習に関して、講座3-5において説明します。

クラウド（サーバ）へ蓄積された情報と現在のセンサーデータを組み合わせた分析例［天候の将来予測］

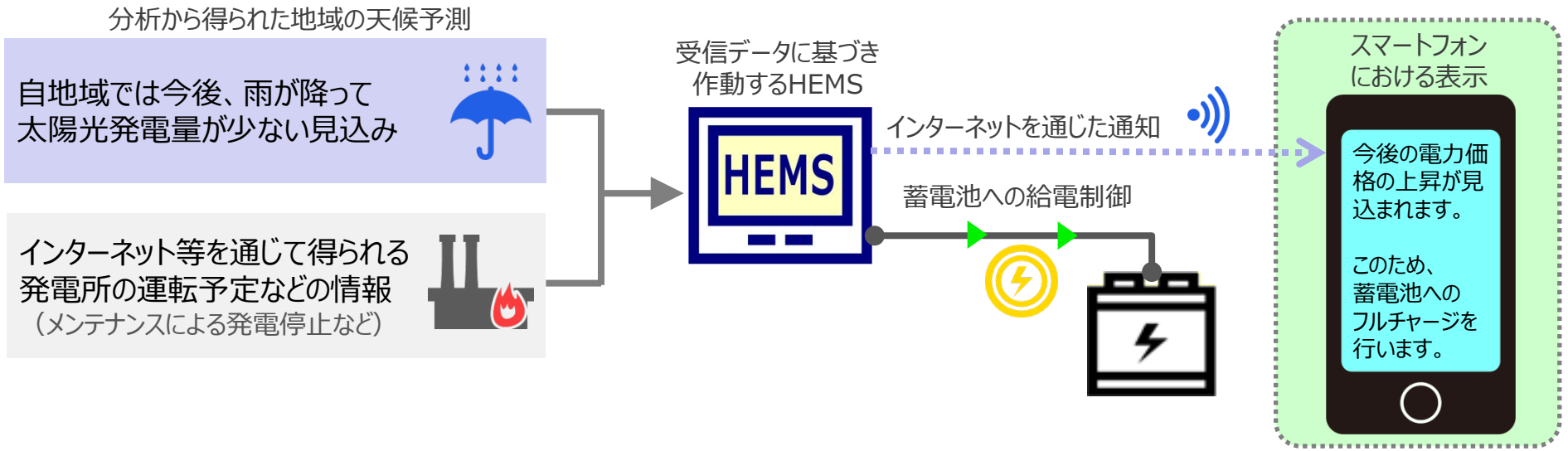


4段階目：分析・データに基づく作動

◆受信したデータを活用して機器が作動したり、人間の行動が変化したりします。

- 分析結果や受信したデータを有効活用して、自動的に機器が作動したり、人間の行動が変化したりします。
  - 受信した情報に応じて作動する機器を総称して、アクチュエーター（actuator）といいます。
  - 自動的な機器の作動としては、スマートフォンへの通知、エアコンの自動的な温度調節が挙げられます。
  - 人間の行動の変化としては、スマートフォンの通知に基づく健康管理、店舗の販売データの分析結果に基づいた仕入れ内容の調整が挙げられます。
- 人間の判断や行動が介在せず、受信データ・分析結果から**自動でプログラムや機械が作動**するケースもあります。
  - コンビニエンスストアでは、販売データや需要予測に基づいて、プログラムが次期の仕入れを決定する「自動発注」が導入されているケースもあります。
- 受信したデータや分析結果に応じてプログラムや機械が作動するケースとして、動作状況のスマートフォンへの**通知**、機械やロボットの**制御**が挙げられます。
  - 講座1-4で示すように、ロボットに関するネットワーク連携が進展することで、ロボットはIoTの一つとして考えられるようになってきました。

受信したデータに基づいて作動する機器の例［HEMSによる蓄電池への給電制御とスマートフォンへの通知］

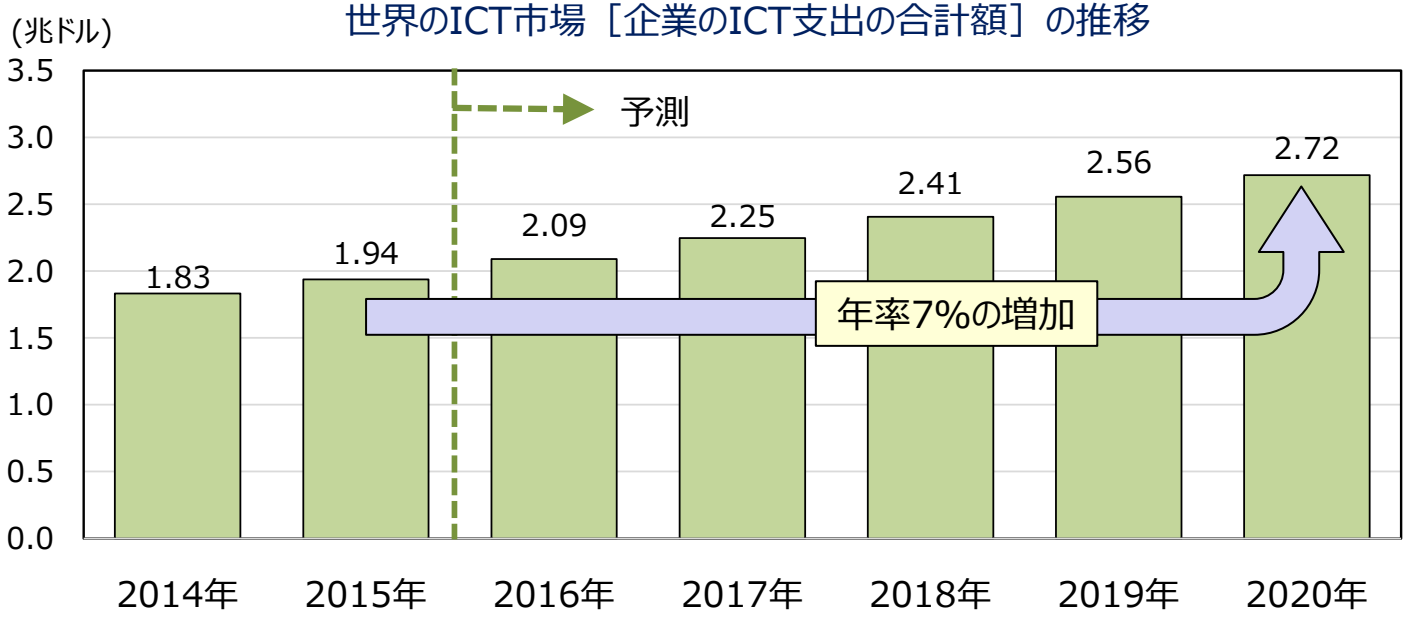


世界のICT市場規模の予測

◆世界のICT市場の市場規模［企業のICT支出の合計額］は、年々増加しています。

- IHS Technology社による推計によれば、世界のICT（情報通信技術）市場の市場規模［企業のICT支出の合計額］は、2015年時点で1.94兆ドルに達していました。
  - IHS Technologyは世界的な市場調査会社であり、2016年に金融情報会社Markitと合併し、IHS Markitとなりました。
  - IHS Technologyの推計において、ICT支出は「企業のICT関連機器、ソフトウェア、サービスへの支出の合計」と定義されています。
- 2020年おける世界のICT市場の市場規模の予測値は2.72兆ドルであり、2015年から年率約7%の支出額の増加が見込まれています。

世界のICT市場［企業のICT支出の合計額］の推移



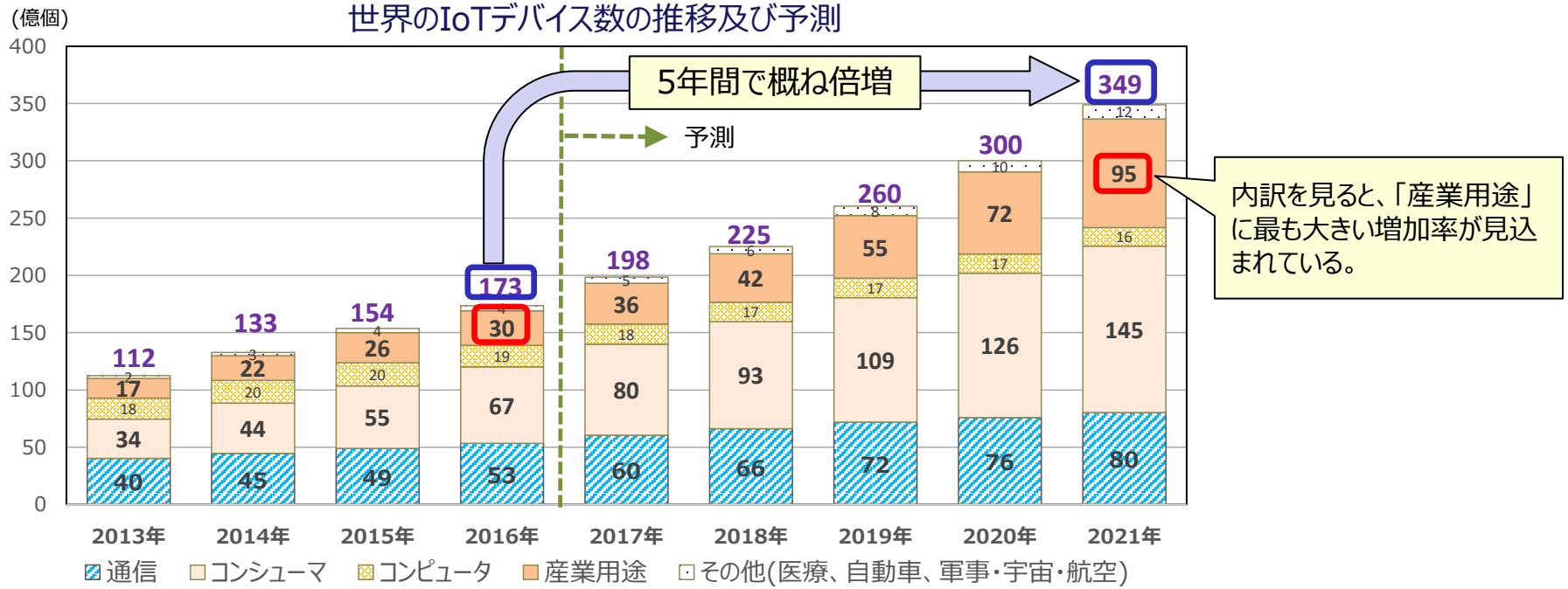
【出所】平成28年版 情報通信白書［総務省］（原典：IHS Technology）に基づき作成
http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h28/html/nc122100.html

□ 世界の企業におけるICT支出の予測値からも、ICTの利活用には大きな成長が見込まれていることが把握できます。

世界のIoTデバイス数の予測

◆IoT機器は世界規模で急速に普及することが予測されています。

- IHS Technology社による推計によれば、2016年時点で世界でIoTデバイス数は173億個でしたが、2021年に349億個へと、5年間で概ね倍増すると見込まれています。
  - IHS Technologyの推計におけるIoTデバイスとは「固有のIPアドレスを持ち、インターネットに接続が可能な機器」を指しています。
- IoTデバイス数の予測の産業別内訳を見ると、「産業用途」の伸びが最も大きく、2016年から2021年の5年間で30億個から95億個へと3倍超の増加が見込まれています。



【出所】平成29年版 情報通信白書［総務省］（原典：IHS Technology）に基づき作成
http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/html/nc133100.html

◘ IoTデバイス数の活用は、特に産業用途において大きな伸びが予測されています。

ドイツ発祥の「Industrie 4.0」と「第四次産業革命」

◆第四次産業革命は、ドイツ政府が公表したIndustrie 4.0が先駆けとなっています。

- 現在は**第四次産業革命**の時期に当たるとされており、この言葉はドイツにおける**Industrie（インダストリー）4.0**が、先駆けとなっています。
  - Industrie 4.0は、2010年にドイツ政府が公表した「High-Tech Strategy 2020 Action Plan」から、議論が始まりました。
  - ドイツにおけるIndustrie 4.0は、「スマートファクトリー」に代表される効率的な工業生産を中心とする概念ですが、「第四次産業革命」は、非製造業を含めた広範な生産性革命を意味しているケースもあります。

- ドイツ経済エネルギー省では、「Plattform Industrie 4.0」を組織して、ウェブサイトにおいて英語とドイツ語での情報発信をしています。

【出所】Plattform Industrie 4.0（英語版）［ドイツ経済エネルギー省］
https://www.plattform-i40.de/I40/Navigation/EN/Home/home.html

- 第四次産業革命は、「蒸気機関」「電力・石油」「電子機器・PC」の、それぞれが牽引した過去の産業革命に続いて**「IoT・クラウドなどによる自律的な最適化」**が牽引するとされています。

各産業革命を牽引する技術と発生時期

	第一次産業革命	第二次産業革命	第三次産業革命	第四次産業革命
牽引する技術	蒸気機関	電力・石油	電子機器・PC	IoT・クラウド
世界における発生時期	18～19世紀初頭	19世紀後半	20世紀後半	21世紀

自律的な最適化が可能となる。

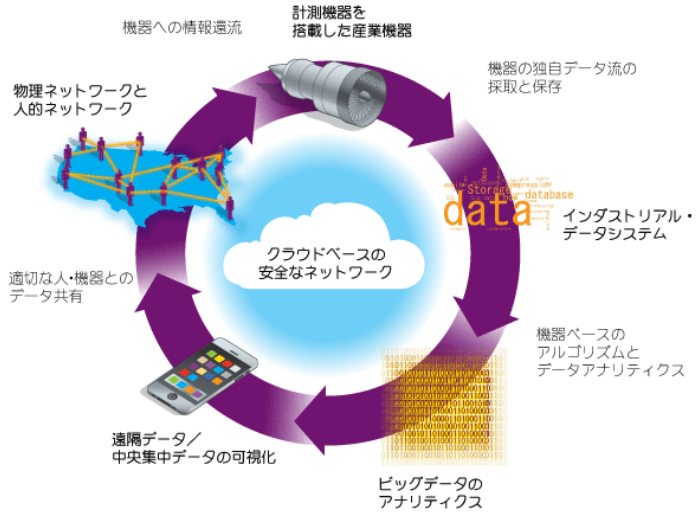
【出所】平成29年版 情報通信白書［総務省］等に基づき作成
http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/html/nc131100.html

アメリカにおける「Industrial Internet」

◆アメリカではIndustrial Internetとして、ICTの産業での活用を推進しています。

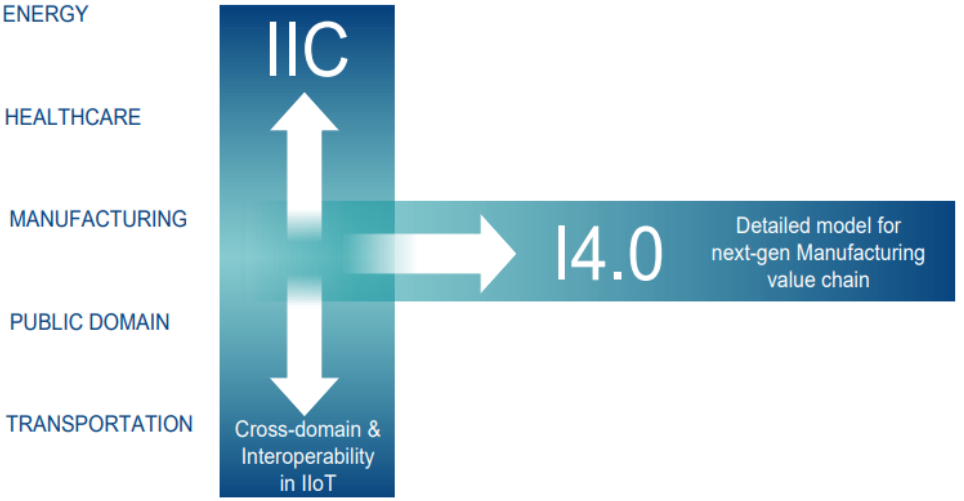
- アメリカでは、GE（ゼネラル・エレクトリック）社が中心となって、**Industrial Internet（インダストリアル インターネット）**と呼ばれる枠組みで、産業におけるインターネット利用を推進し、生産性向上を目指しています。
  - アメリカにおけるIndustrial Internetでは、エネルギー、ヘルスケア、製造業、公共、運輸の5領域を主な対象としています。
  - Industrial Internetが製造業を対象としている点はIndustrie 4.0と共通していますが、製品販売後のデータ収集やデータ活用に重点をおいています。
- Industrial Internetを推進するために、2014年、アメリカに本部を持つ**Industrial Internet Consortium**（略称：IIC）が設立され、世界のICT関連企業が会員となっています。
  - IICの会員企業の中には、ドイツ企業や日本企業もあり、2018年3月末時点で230社以上の企業が会員となっています。
  - IICは、ドイツのPlattform Industrie 4.0と情報交換や共同作業を行っており、共同執筆の報告書を公開しています。

インダストリアル インターネットにおけるデータのループ



【出所】インダストリアル・インターネット［GE（ゼネラル・エレクトリック）］
https://www.ge.com/jp/industrial-internet

5領域を対象とするIICと製造業に注力するIndustrie 4.0のイメージ



【出所】Industrial Internet Consortium
https://www.iiconsortium.org/iic-i40-joint-work.htm

中国における「中国製造2025」と「インターネットプラス」

◆中国は「中国製造2025」「インターネットプラス」によって、製造業の強化を推進しています。

- 中国政府は2015年に**中国製造2025（Made in China 2025）**を公表し、中国建国100周年の2049年までに、品質や生産効率を含めた製造業の総合力において、世界の先頭グループに入る目標を掲げています。
  - 2025年と2035年において中間目標が設定されており、最初の10年の行動綱領として「中国製造2025」と命名されています。
- 中国製造2025では、9つの戦略任務と10の重点分野が示されています。

「中国製造2025」における9つの戦略任務と10の重点分野

戦略任務
- （1）国家の製造業イノベーション能力の向上
- （2）情報化と産業化のさらなる融合
- （3）産業の基礎能力の強化
- （4）品質・ブランド力の強化
- （5）グリーン製造の全面的推進
- （6）**重点分野**における飛躍的発展の実現
- （7）製造業の構造調整のさらなる推進
- （8）サービス型製造と生産者向けサービス業の発展促進
- （9）製造業の国際化発展レベルの向上

→ **重点分野**
- ① 次世代情報通信技術
- ② 先端デジタル制御工作機械とロボット
- ③ 航空・宇宙設備
- ④ 海洋建設機械・ハイテク船舶
- ⑤ 先進軌道交通設備
- ⑥ 省エネ・新エネルギー自動車
- ⑦ 電力設備
- ⑧ 農業用機械設備
- ⑨ 新材料
- ⑩ バイオ医薬・高性能医療機械

【出所】「中国製造 2025」の公布に関する国務院の通知の全訳［科学技術振興機構］
https://www.jst.go.jp/crds/pdf/2015/FU/CN20150725.pdf

- 中国政府は**インターネットプラス**と呼ばれるインターネットを活用した産業の競争力強化方針を示しており、「中国製造2025」と「インターネットプラス」を統合した政策を推進しています。

【出所】'Made in China 2025' gets boost from innovation-driven strategy［中華人民共和国 国務院］
http://english.gov.cn/premier/news/2018/02/24/content_281476057232092.htm

日本における「Connected Industries」

◆日本政府は、目指すべき産業の姿として「Connected Industries」を提唱しています。

- 2017年3月にドイツのハノーバーにおいて、日本政府とドイツ政府の間で第四次産業革命に関する**日独協力の枠組み**を定めた**「ハノーバー宣言」**が署名されました。
  - ハノーバー宣言では、「サイバーセキュリティ」「国際標準化」「規制改革」「中小企業支援」「研究開発」「プラットフォーム」「デジタル人材育成」「自動車産業」「情報通信分野の協力」に関して日独間の協力が合意されました。

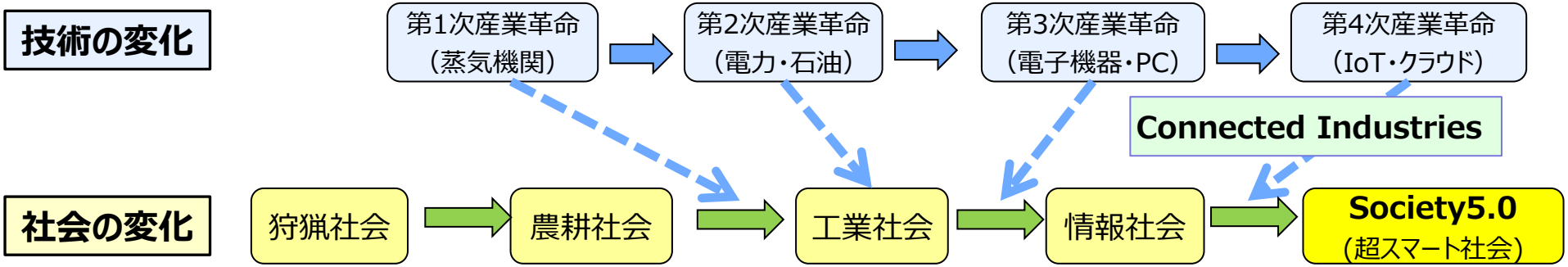
【出所】ハノーバー宣言［経済産業省］http://www.meti.go.jp/press/2016/03/20170320002/20170320001.html

- 「ハノーバー宣言」に合わせて、日本が目指すべき産業の在り方として**「Connected Industries（コネクティッド インダストリーズ）」**の概念が示されました。
  - Connected Industriesは、「人、モノ、技術、組織等が様々につながることにより、新たな価値創出や競争力の強化をもたらす」という考え方です。

【出所】Connected Industries［経済産業省］http://www.meti.go.jp/policy/mono_info_service/connected_industries/

- 各産業革命における「技術の変化」は「社会の変化」をもたらし、第四次産業革命にて実現されるConnected Industriesは、新たな社会である**Society（ソサエティー）5.0**につながるとされています。

「技術の変化」と「社会の変化」の関係



【出所】新産業構造ビジョン［経済産業省］に基づき作成
http://www.meti.go.jp/press/2017/05/20170530007/20170530007-2.pdf

Society（ソサエティー）5.0

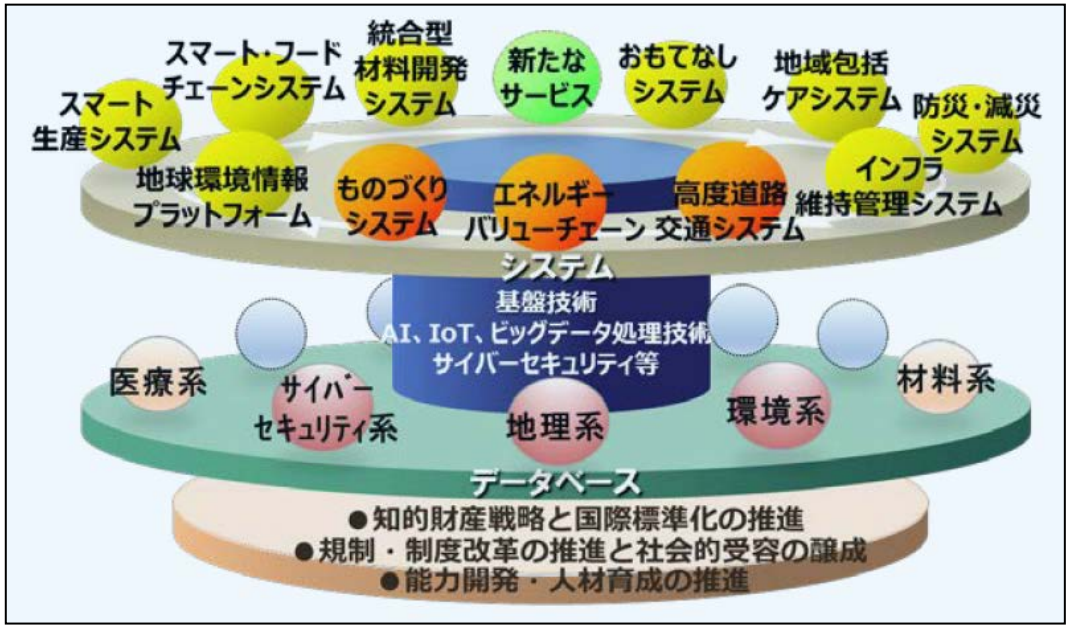
◆日本政府はIoTやデータを活用することで、Society5.0の実現を目指しています。

- **Society（ソサエティー）5.0**は、日本政府が目指す**「質の高い生活ができる人間中心の社会」**です。
  - Society 5.0は、2016年1月に閣議決定された第5期科学技術基本計画にて示されました。
  - 「5.0」には狩猟社会、農耕社会、工業社会、情報社会に続く、5番目の経済社会という意味があります。

【出所】科学技術基本計画［内閣府］http://www8.cao.go.jp/cstp/kihonkeikaku/index5.html

- **Society 5.0は、IoTやデータを活用**することで、製造業（ものづくり）、農業、交通、サービス業等の**広範囲の課題解決によって実現**されます。

Society 5.0を実現するプラットフォームのイメージ



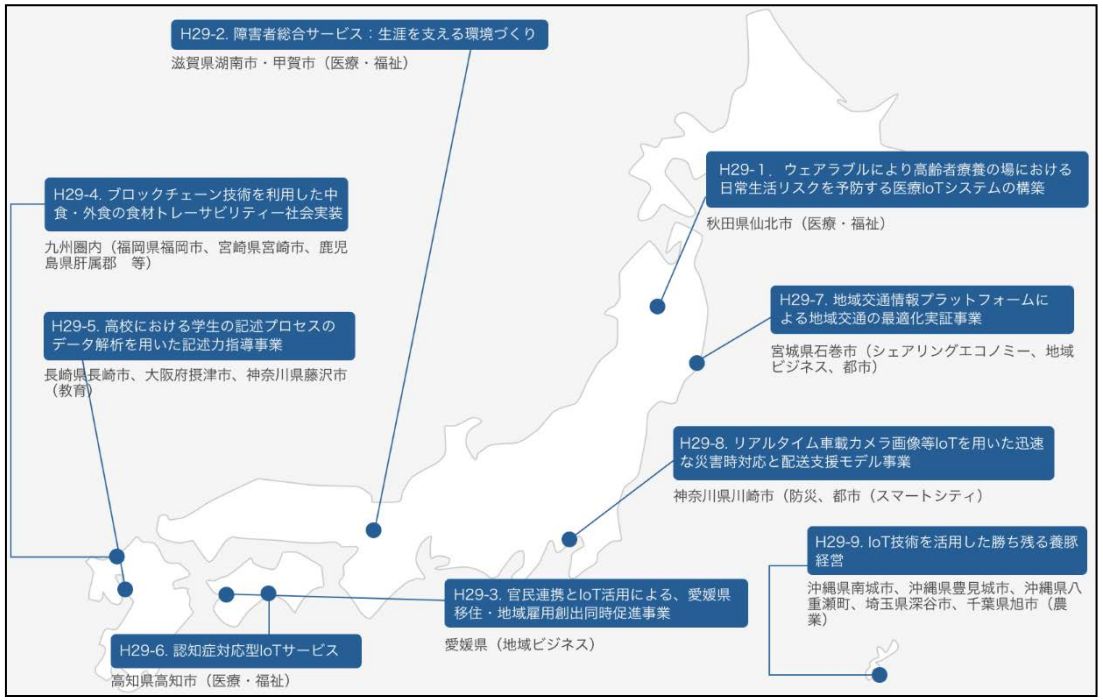
【出所】科学技術イノベーション総合戦略2017（概要）［内閣府］
http://www8.cao.go.jp/cstp/sogosenryaku/2017.html

IoTの活用の実証事業の公募

◆総務省、経済産業省では、IoTの利活用に関する実証事業を公募しています。

- 総務省では、平成28年度よりIoTサービスの実証実験およびサービス展開時に必要なルールの明確化を行う**「IoTサービス創出支援事業」**の公募を行い、その成果等を**「身近なIoTプロジェクト」**にて紹介しています。
  - 平成29年度の「総務省IoTサービス創出支援事業」において、125件の提案から下の地図に示す9件が採択されました。

総務省IoTサービス創出支援事業（平成29年予算）に採択された9事業および「身近なIoTプロジェクト」



身近なIoTプロジェクト
midika-iot.jp

【出所】身近なIoTプロジェクト［総務省］
https://www.midika-iot.jp/project/

- 経済産業省では、平成27年度よりIoT等を用いて、訪日外国人旅行者の情報を本人の同意のもとで利活用する**「おもてなしプラットフォーム」**の実証事業を公募しています。

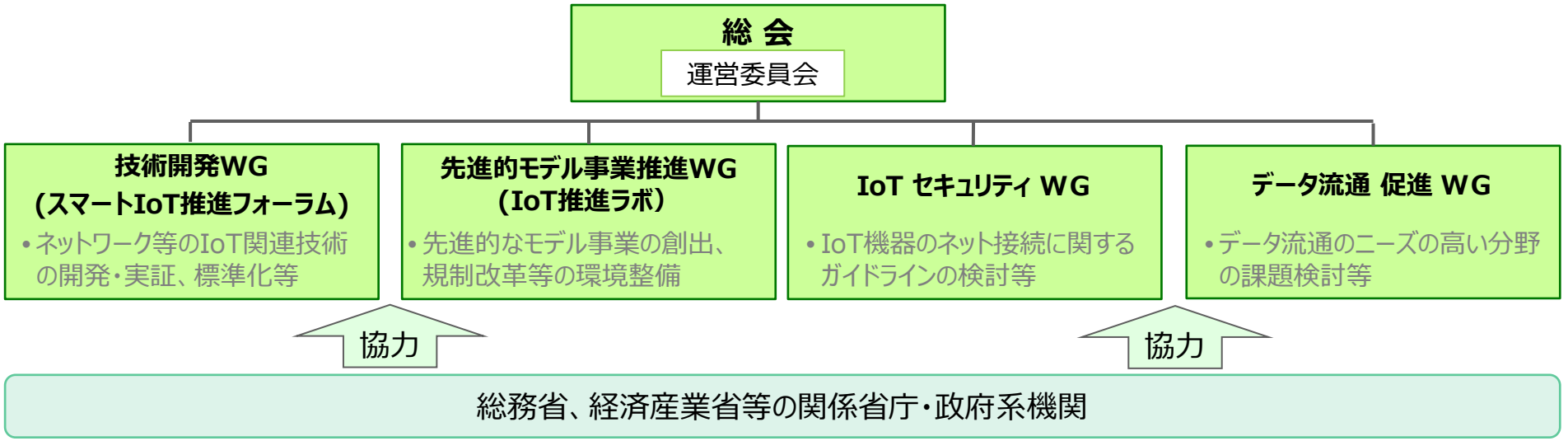
【出所】「おもてなしプラットフォーム」の平成29年度実証事業を実施［経済産業省］
http://www.meti.go.jp/press/2017/09/20170929003/20170929003.html

IoT推進コンソーシアム

◆2015年、IoTの利活用を産学官で推進する「IoT推進コンソーシアム」が設立されました。

- 2015年に設立された**IoT推進コンソーシアム**では、総務省・経済産業省などの省庁が協力する形で**産官学が一体となってIoT関連政策、事業**を行っています。
  - 技術開発や標準化、規制改革などの環境整備、セキュリティの検討、データ流通のそれぞれでワーキンググループが構築されています。
  - 2018年3月13日時点で法人会員が3,513社、特別会員として地方自治体が59団体、6省庁が参加しています。

IoT推進コンソーシアムの組織構造と協力体制（2017年9月時点）



- IoT推進コンソーシアムは、2016年10月にアメリカに本部を持つ**IIC（Industrial Internet Consortium）との間で「日米IoT分野の協力に係る覚書」**への署名を行いました。
  - 覚書では、「グッドプラクティスの発掘・共有」「テストベッドや研究プロジェクトの協力」「アーキテクチャ等の相互運用性の確保」「標準化に関する協力」といった協力方法が示されています。

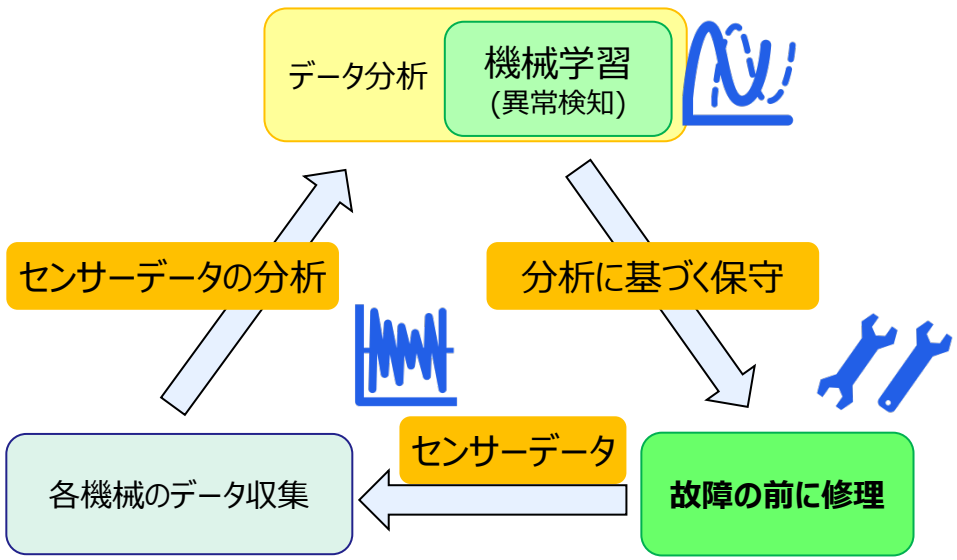
【出所】2017年度（第3回）総会の公開資料［IoT推進コンソーシアム］に基づき作成 http://www.iotac.jp/event/plenary2017/

製造業におけるデータの利活用分野

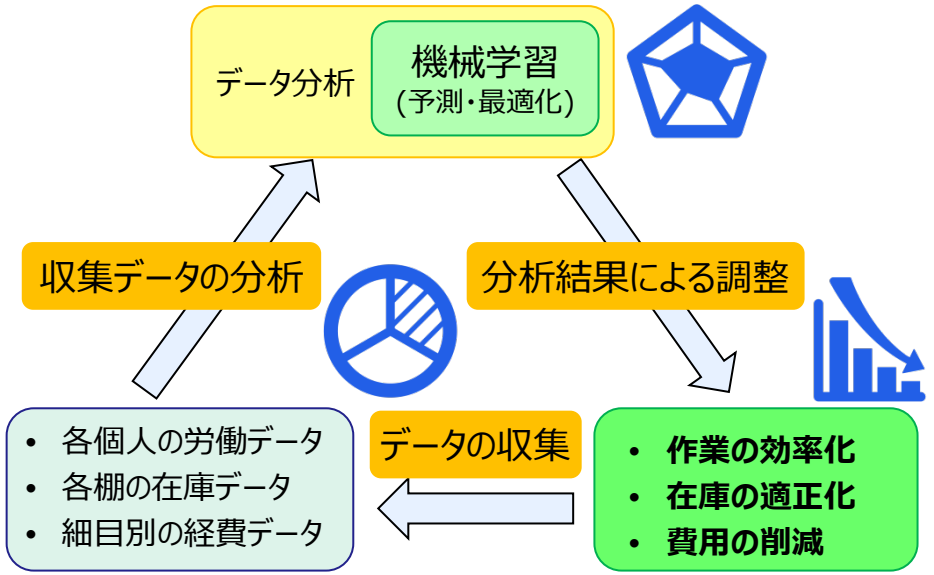
◆製造業に着目しても、様々な業務でIoT･データの利活用が可能です。

- 製造面：製造用機械の動作と不良品の発生データをセンサーから読み取り、製造用機械の**保守･メンテナンスの時期や内容を最適化することで**、**故障の前に修理**できます。
- 在庫管理面：在庫情報･入出庫情報のデータを収集･分析することで、**作業効率や在庫効率を改善し、費用を削減**することができます。

製造面：IoTによる製造機械の異常検知



在庫管理面：IoTによる在庫管理



□ IoT･データを活用すると、全体の傾向把握のみならず、個々の機械、商品に応じた個別対応が可能になります。

様々な産業におけるIoT･データの活用（X-Tech）

◆IoT･データの利活用は金融、医療といった様々な分野で考えられます。

- IoTをはじめとする様々な新技術を活用した新サービスを総称して、**X-Tech**（エックステック）と呼びます。
  - このXの部分が産業毎に変化し、金融なら「FinTech」、医療なら「MediTech」と呼ばれます。
  - ICT関連用語では、様々な文字が入る部分に「x」を入れることで、総称として示す場合があります。他の事例として、講座1-3に示すxR(エックスアール)、講座2-2に示すXaaS（ザース）が挙げられます。

- **FinTech（フィンテック）**と呼ばれる金融分野での活用では、金融機関の利用や購買のデータから、家計簿・資産状況を示すアプリが挙げられます。
  - レシート（領収書）をスマートフォンのカメラで撮影するだけで、購買情報を記録できるFinTechアプリもあります。

家計簿アプリ（マネーフォワード）

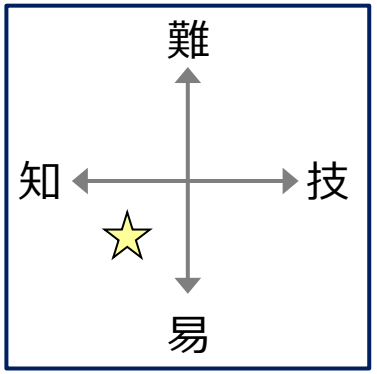
【出所】マネーフォワード［株式会社マネーフォワード］
https://moneyforward.com/

- **MediTech（メディテック）**と呼ばれる医療分野での活用では、生活習慣や体重等のデータから、健康管理を支援するアプリが挙げられます。
  - 体重計等が収集した情報を、無線でスマートフォンやPCに送信できるMediTechアプリもあります。

健康管理アプリ（ヘルスプラネット）と対応する体重計

【出所】ヘルスプラネット［株式会社タニタヘルスリンク］
http://www.healthplanet.jp/

総務省　ICTスキル総合習得教材



［コース1］データ収集

1-2：データ収集技術とウェアラブルデバイス

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_1_2.pdf

	1	2	3	4	5
［コース1］データ収集		●			
［コース2］データ蓄積					
［コース3］データ分析					
［コース4］データ利活用					

本講座の学習内容［1-2：データ収集技術とウェアラブルデバイス］

【講座概要】

- 人工的な記録を読みとるリーディング技術と先端事例として、RFIDを説明します。
- IoTに利用される無線通信を概観し、長距離通信ができるLPWAの利用例を紹介します。
- 自然の情報を読み取るセンシング技術とその利用例を紹介します。
- センシング技術の応用例として、身につけるウェアラブルデバイスとその活用事例を紹介します。

【講座構成】

座学

- [1] リーディング技術とRFID
- [2] IoTと無線通信
- [3] センシング技術とその利用例
- [4] ウェアラブルデバイスの活用例

【学習のゴール】

- ✓ IoTに関わるリーディング技術として、普及しているRFIDの利活用を把握する。
- ✓ 長距離通信ができるLPWAの活用例を把握する。
- ✓ 様々なセンシング技術の種類と対象情報、利活用例を把握する。
- ✓ ウェアラブルデバイスの活用事例を紹介できる。

リーディングとセンシング

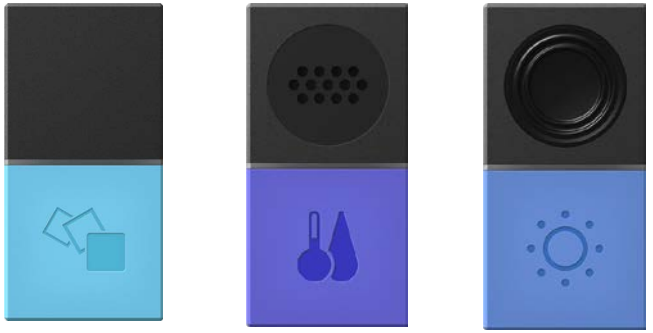
◆リーディングは人工的な記録を読み取る一方で、センシングは自然の情報を感知します。

- **リーディング**（Reading：読み取り）は**人工的な記録情報を読み取る**ことを指し、読み取る機器は**リーダー**です。
  - 具体的な機器としては、バーコードリーダーやSDカードリーダーが該当します。
- **センシング**（Sensing：感知）は**自然の情報を読み取る**ことを指し、感知する機器は**センサー**です。
  - リーダーとセンサーは、情報を読み取る点や情報収集を担う点では共通しています。
- IoT機器がセンサーになることで、人間の五感のように、光、音、臭い、味、温度、圧力などの情報収集をすることができます。
  - ソニー株式会社が発売しているMESH（メッシュ）は、動き、温度・湿度、明るさのセンサーとなっているものがあり、それらの変化を感知すると、無線でスマートフォン等に通知します。

バーコードリーダー
［バーコードのリーディング］

メディアリーダー
［SDカード等のリーディング］

MESH（メッシュ）
［動き、温度・湿度、明るさのセンシング］



【出所】MESH［ソニー株式会社］
http://meshprj.com/jp/

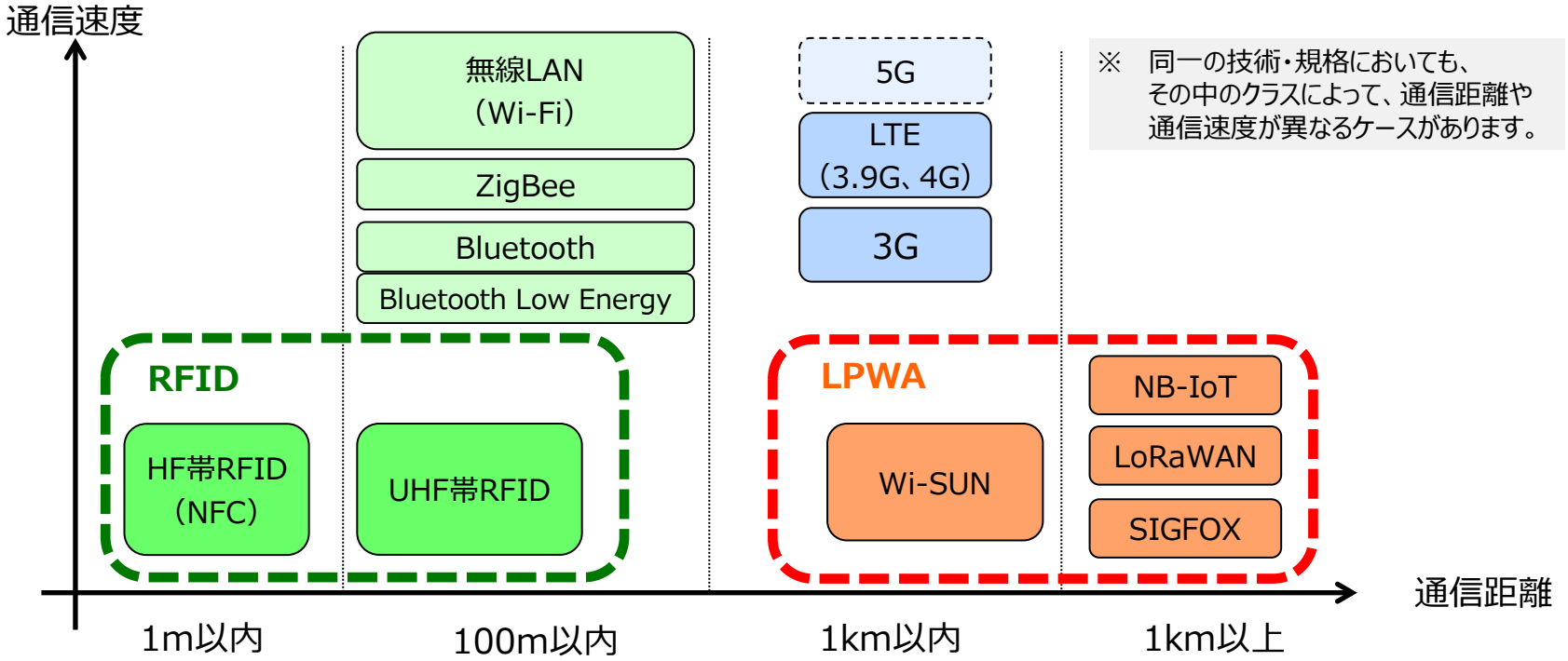
□ この講座では、リーディングに関する内容を説明した後、センシングに関する内容を説明します。

IoTで注目される通信技術としてのRFID

◆IoTの普及に伴って、RFIDやLPWAと呼ばれる通信技術が注目されています。

- IoTと親和性が高い通信技術として、**RFID**（アールエフアイディー：Radio Frequency Identifier）と**LPWA**（エルピーダブリューエー：Low Power, Wide Area）が挙げられます。
  - RFIDおよびLPWAは通信速度は低速ですが、省電力性等から小規模かつ長期間の情報の送受信に適しています。
  - RFID、LPWAのそれぞれ技術の総称であり、それぞれの技術の中には個別の規格があります。
- RFIDは、人工的な情報を読みとるリーディング技術であるとともに、接触なく情報を読む無線通信技術でもあります。

主な無線通信技術・規格における通信距離と通信速度の目安



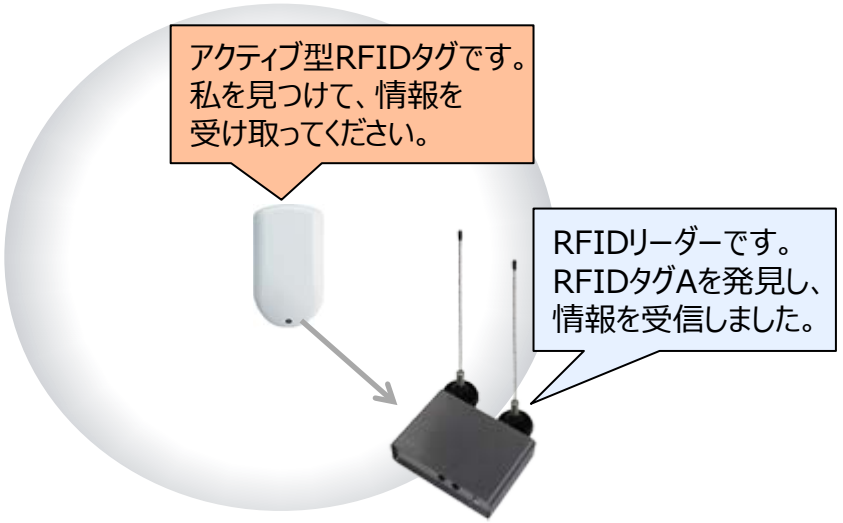
【出所】第４次産業革命における産業構造分析とIoT・AI等の進展に係る現状及び課題に関する調査研究
[総務省（調査委託先：三菱総合研究所）] 図表3-8等に基づき作成
http://www.soumu.go.jp/johotsusintokei/linkdata/h29_03_houkoku.pdf

IoTとの親和性が高いリーディング技術 RFID

◆電波を利用する認証技術のRFIDは、アクティブ型とパッシブ型に大別できます。

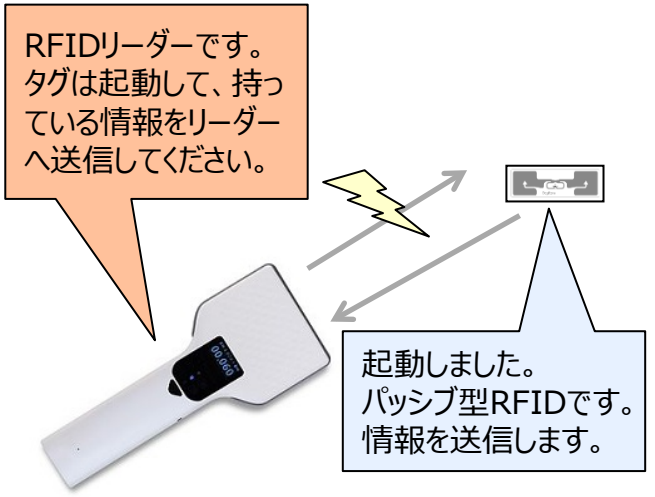
- **RFID**は、接続された**RFIDタグ**の情報を短距離無線通信（数cm～数十m）によって送受信するシステムで、**電波を利用する認証技術の総称**です。
  - RFIDはデータが書き込まれるタグ（札）は、RFIDタグ、ICタグ、電子タグなどとも呼ばれますが、この講座ではRFIDタグと呼びます。
- RFIDタグには**電池が必要なアクティブ型**、リーダーの電磁波を動力源とする**紙状・電池不要のパッシブ型**があります。
  - アクティブ型RFIDは電池をとりつけ、相対的に高出力の通信が可能となるため、通信距離がパッシブ型RFIDに比べて長い傾向にあります。
- パッシブ型RFIDタグには、紙状・電池不要という特性があり、特にIoTとの広範な連携が期待されています。
  - パッシブ型RFIDタグは、商品の値札に埋め込んでもほとんど厚みは増さず、光に透かしてようやく確認することができます。

アクティブ型RFID



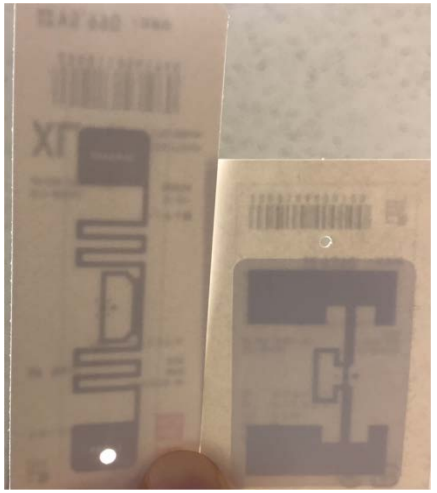
【出所】吉川工業株式会社
http://www.ykc.co.jp/product_technology/rfid/active_rfid.html

パッシブ型RFID



【出所】株式会社東北システムズ・サポート
https://rfid.tss21.co.jp/product/swing-u/

商品の値札に埋め込まれた
パッシブ型RFIDタグ



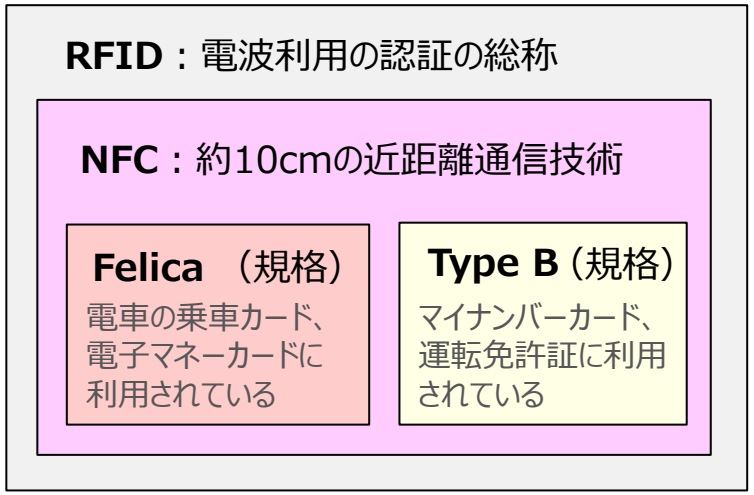
【出所】株式会社
ファーストリテイリング

RFIDにおけるNFCとFelica

◆RFIDは、交通機関の乗車カード、電子マネーカードとして幅広く普及しています。

- RFIDには、**NFC**（エヌエフシー：Near Field Communication）という通信距離が約10cmの技術があります。
  - NFCの通信距離が短いHF（High Frequency：短波）の通信技術ですが、数十mまで通信できるUHF（Ultra High Frequency：極超短波）の通信技術もあります。
- NFCの規格の一つである**Felica**（フェリカ）は、電車の乗車カード（例：Suica）や電子マネー（例：楽天Edy）などの形で広く利用されています。
  - NFCには、Felica以外にも規格があり、マイナンバーカードや運転免許証に採用されている規格は、通称「Type B」と呼ばれています。
- FeliCa規格のRFIDタグを搭載することで、リーダーにかざすだけで電子マネーが利用できるスマートフォンもあります。
  - 金属以外の物であれば、RFIDタグとリーダの間に存在しても、ほとんど影響を受けずにデータの読み書きができます。

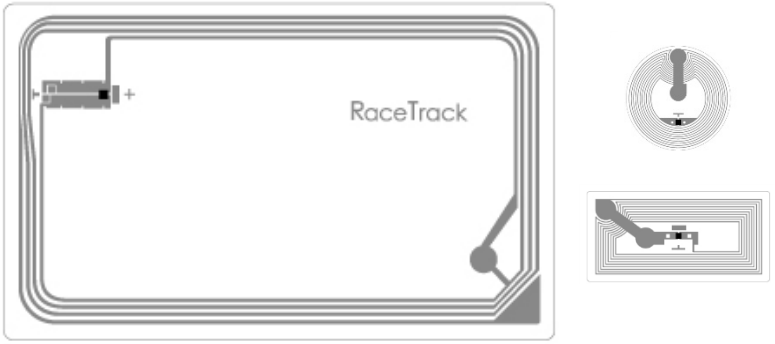
RFID・NFC・Felicaの包含関係



NFCのロゴ

【出所】NFC FORUM
https://nfc-forum.org/

NFCのRFIDタグ（シール）の例



【出所】東北システムズ・サポート
https://www.tss21.co.jp/product/rfid/

パッシブ型RFIDの特徴と低価格化・普及の推進

◆パッシブ型RFIDは、価格低下と普及に関する官民の取り組みが実施されています。

- パッシブ型RFIDとバーコードやQRコードを比較すると、長所の方が数多くありますが、短所として**相対的に作成単価が高い**ことが挙げられます。
  - RFIDタグは小さくとも電子部品ですが、バーコードとQRコードは、紙と印刷代だけで安価に作成することができます。

パッシブ型RFIDタグをバーコード・QRコードと比較した場合の長所と短所

【長所】
- ○ **規格に応じて、最大数十mまで読みとれる**
- ○ **複数のタグを一括して読みとることができる**
- ○ **情報を読むだけではなく、書き込むこともできる**
- ○ **箱の中等に入っていても読み取れる**
- ○ **表面が汚れていても読み取れる**

【短所】
- × **作成単価が相対的に高い**

RFIDタグ

バーコード

QRコード

- こうしたパッシブ型RFIDの短所を改善し、それによって更なる普及と活用を目指す取り組みも進められています。
- 大日本印刷株式会社は2017年3月において、単価10円台のRFIDを2020年までに5円以下、2025年に1円を目指す開発に着手することを発表しました。

【出所】IoTの拡大に向けてコンビニ向けの低価格ICタグの開発に着手［大日本印刷株式会社］
http://www.dnp.co.jp/news/10132735_2482.html

- 2017年4月、経済産業省では、**「コンビニ電子タグ1000億枚宣言」**を策定し、5系列のコンビニエンスストアの全ての取扱商品に2025年までにRFIDタグを貼りつけ、商品の個品管理を実現する方針をとりまとめました。

【出所】「コンビニ電子タグ1000億枚宣言」を策定しました［経済産業省］
http://www.meti.go.jp/press/2017/04/20170418005/20170418005.html

パッシブ型RFIDの実用化例

◆パッシブ型RFIDは、既に様々なサービスにおいて実用化されています。

- 2017年11月、株式会社ファーストリテイリングはユニクロの全ての店舗の衣料品に、精算が瞬時に可能になるRFIDタグを1年以内に取り付ける方針を明らかにしました。
  - セルフレジでは購入予定の商品をRFIDリーダーが読み取れる範囲に置くことで、複数の商品のタグをまとめて読みとって金額を算出します。
  - 株式会社ファーストリテイリングによる実証結果では、セルフレジは有人レジと比較して、精算所要時間が最大で約3分の1となると示されました。
- 東京都港区図書館では、図書の貸出管理をRFIDで行っており、本を閉じたまま一括して認識することができます。
- 回転寿司店においては、皿にRFIDタグを取り付けることで、会計時の皿の分類や数え上げの手間を解消しています。

RFIDを利用した衣料品店のセルフレジ

【出所】株式会社ファーストリテイリング

RFIDを利用した図書貸出機

【出所】東京都港区図書館

RFIDを利用した回転寿司店の会計処理

【出所】北一食品株式会社

□ RFIDタグが使い捨てできるレベルにまで低価格化すれば、製品管理や会計処理が大きく変わると考えられます。

IoTに利用される距離100m以内の通信技術

◆Wi-FiやBluetoothといった距離100m以内の無線通信もIoTに活用されています。

- 一般に100m以下の通信距離となる無線通信として、無線LANの代表的な規格である**Wi-Fi**（ワイファイ）に加えて、**Bluetooth**（ブルートゥース）、**Zigbee**（ジグビー）が挙げられます。
  - Bluetoothは、PC用の無線キーボード・マウスやスマートフォン等で利用できる無線ヘッドフォンとして、幅広く利用されています。
  - Zigbeeは最大65,536台の同時接続が可能であり、多数のセンサーから同時に情報収集することに適しています。

Wi-Fiの認定製品のロゴ

【出所】Wi-Fi Alliance
https://www.wi-fi.org/ja

Bluetoothのロゴ

【出所】Bluetooth SIG
https://www.bluetooth.com/ja-jp

Zigbeeのロゴ

【出所】Zigbee Alliance
http://www.zigbee.org/

- 無線LANを利用できる端末は単体でインターネットに接続することができますが、Bluetooth、Zigbeeは利用する端末は、一般に直接インターネットに接続することはできません。
  - LANが利用できる端末には、ルーターと呼ばれる機器からインターネット通信における住所に相当するIPアドレスが与えられます。

- 小型IoT端末は、Bluetoothやその低消費電力の規格の**BLE**（Bluetooth Low Energy）等の無線通信で別のインターネット接続機器（スマートフォンやPC）を経由して、インターネットとつながることが一般的です。
  - BLEは、その正式名称の通り、従来のBluetoothに比べて低消費電力であることが長所であり、使い方によってはボタン電池一つで1年以上継続稼働するとされています。

□ IoT（インターネットオブシングス）とは言っても、端末とインターネットとの接続は間接的であるケースもあります。

IoTに利用される距離100m以上の通信技術（LPWA）

◆LPWAは消費電力が小さく、距離100mを超える通信が可能な技術です。

- **LPWA**は「**L**ow **P**ower, **W**ide **A**rea」に由来する**電力消費が小さく100m以上の通信ができる技術**です。
- LPWAの代表的な規格として**Wi-SUN**(ワイサン)、**NB-IoT**（エヌビーアイオーティー）、**LoRaWAN**（ローラワン）、**Sigfox**（シグフォックス）が挙げられます。

Wi-SUNのロゴ

【出所】Wi-SUN Alliance
https://www.wi-sun.org/

NB-IoTのロゴ

【出所】GSMA
https://www.gsma.com/iot/nb-iot-logo-download-pack/

LoRaWANのロゴ

【出所】LoRa Alliance
https://www.lora-alliance.org/

Sigfoxのロゴ

【出所】Zigbee Alliance
https://www.sigfox.com/en

- Wi-SUNは「**Wi**reless **S**mart **U**tility **N**etwork」に由来し、LPWAでは相対的に速い通信速度を持ちます。
  - Wi-SUNの通信速度は200kbpsとLPWAの中では高速ですが、最大通信距離はLPWAの中では相対的に短く1km未満となっています。最大通信距離が相対的に短いことからWi-SUNをLPWAに含めて分類しないケースもあります。
- NB-IoT、LoRaWAN、Sigfoxのそれぞれの通信速度は、Wi-SUNよりも低速（100kbps以下）ですが、最大通信距離は1km以上となっています。
  - NB-IoTのNBは、「**N**arrow **B**and」の略で、モバイル通信の国際標準化を推進する3GPP(3rd Generation Partnership Project)の仕様です。
  - LoRaWANのLoRaは、「Long Range」に由来し、アメリカを中心とする非営利団体のLoRaアライアンスが推進しています。
  - Sigfoxはフランスの通信事業者「SIGFOX」が国際的に展開しており、通信速度は100bpsと低速ですが、通信距離は最大50kmとより広範囲です。

LPWAの実証事例（小1児童登下校お知らせサービス）

◆複数の通信技術の特性を活かす形での実証実験が行われています。

- 静岡県藤枝市とソフトバンク株式会社は、ソフトバンクが市内に敷設したLPWAネットワークを活用した「藤枝市小1児童登下校お知らせサービス実証実験」を、2017年10月から藤枝市内にある一部の小学校で開始しました。
- 登下校のタイミングを昇降口におけるBLE（Bluetooth Low Energy）で把握し、登下校途中の位置情報はLoRaWANにて送信しています。
  - 昇降口への立ち寄りを正確に把握できるBLE、校区全体において位置情報を送信できるLoRaWANを組み合わせた実証実験となっています。

「藤枝市小1児童登下校お知らせサービス実証実験」のイメージ図

【出所】「藤枝市小1児童登下校お知らせサービス実証実験」を開始［ソフトバンク株式会社］
https://www.softbank.jp/corp/group/sbm/news/press/2017/20171016_01/

LPWAの実用化例（離島における水道検針）

◆LPWAの長距離通信は、離島における水道検針で実用化されています。

- 「Sigfox 自動検針コンソーシアム」は、2017年11月に兵庫県姫路市の家島諸島内の水道検針をIoTで行う自動検針システムを本稼働させ、請求データとしての利用を開始しました。
  - 「Sigfox 自動検針コンソーシアム」は、第一環境株式会社・アズビル金門株式会社・KDDI株式会社・京セラコミュニケーションシステム株式会社の4 社で構成されています。
- 家島諸島内28ヵ所で収集された検針データは、Sigfox基地局を通じてデータセンターに集約され、Webサイトから確認できます。
  - SigfoxはLPWAの中でも通信距離が最大50kmと相対的に長く、離れた小島や山間部での通信に活用できます。

自動検針に関する4社の協力体制

Sigfox無線発信機付き水道メーター

【出所】水道向けSigfox自動検針ソリューション［第一環境株式会社］http://www.daiichikankyo.co.jp/news/2517/

□ リーディング・無線通信に関するパートは以上となり、続いてセンシングに関するパートに入ります。

センシングの対象となる情報

◆センシング対象となる自然の情報には様々なケースが考えられます。

- センシングの対象となる情報は大別して下の4種が挙げられ、それぞれのセンシングから得た情報が活用されています。

◆日射量や風のような自然環境に関する情報

◆機械の温度、圧力、電圧といった人工物やその動作状況に関する情報

◆動物の体温、心拍数、血圧といった意識的なコントロールができない身体に関する情報

◆人間の体の動き、発声、押したボタンなどの意識的な行動に関する情報

活用例

自然環境の状況記録と自然に関する分析・予報

異常検知と機械の効率的な稼働に関する分析

健康状態のデータとしての把握、医学における分析

電子機器への入力・操作

風量計

工場内のセンサー・計器

血中酸素濃度計

キーボードのタイピング

五感で感知可能な情報のセンシング

◆人間の五感で感知可能な情報は、センシングによって収集可能です。

- **センサー**を利用すると、光や音の強さなど、**人間の感覚では定量化が難しい情報**でも**定量的に測定**できます。
  - カメラやマイクは、画像、音声の詳細な情報を把握できるセンサーとして考えることができます。

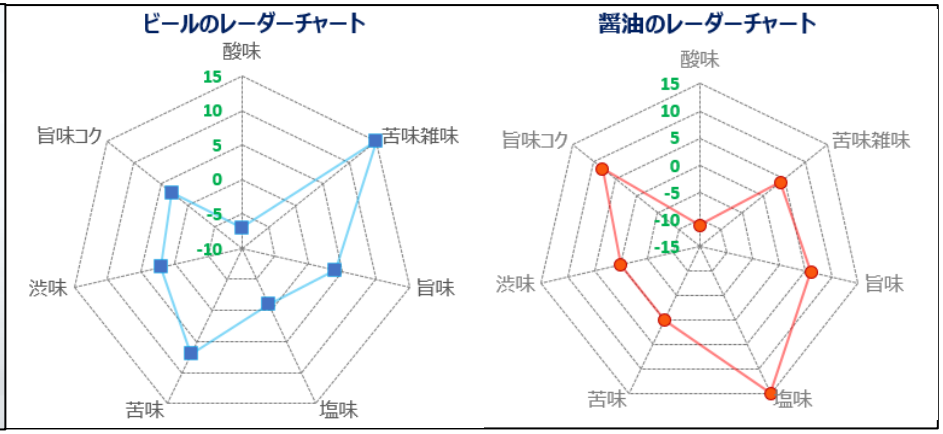
人間の五感で感知できる情報とセンサー

感覚	センサー	用途
視覚	照度センサー、カメラ	農場の日射量管理、小売店における顧客数の把握
聴覚	マイクロフォン	防犯、音声入力
嗅覚	臭気センサー	アルコールや危険物質の検知
味覚	味覚センサー	食品開発、医薬品開発
触覚	圧力センサー、温度センサー	スイッチ、自動的な温湿度の調整

- 株式会社インテリジェントセンサーテクノロジーでは、「世界で初めて味認識装置の実用化に成功した」ことをウェブサイトにて説明しています。
  - 精度の高低はあっても、人間の五感で感知できる情報は、センサーでの情報収集が可能です。

味認識装置TS-5000Z

測定例（ビール、醤油のレーダーチャート）



【出所】株式会社インテリジェントセンサーテクノロジーの資料に基づき作成
http://www.insent.co.jp/products/taste_sensor_index.html

五感を代替するセンシング事例と人工知能

◆視覚情報や聴覚情報のセンシングと人工知能の技術を組み合わせた製品もあります。

- カメラやマイクといった**センサーと人工知能を組み合わせる**ことで、新たな製品・サービスが生まれています。

- フランスのNETATMO社が発売した「NETATMO WELCOME」は、カメラで撮影した写真・動画の情報を人工知能が判別し、必要に応じてインターネットを通じてスマートフォン等へ情報を発信します。
  - 「NETATMO WELCOME」は、顔認識機能を搭載しており、事前に登録していない人が写った場合のみ利用者に通知することが可能です。

- Amazon社が発売した「Amazon Echo」およびGoogle社が発売した「Google Home」は、マイクでのセンシングに加えて、人工知能による音声認識によって声による操作が可能です。
  - 利用者の話し方や声の特徴をあらかじめ、登録しておくことが可能です。

視覚情報を代替するNETATMO WELCOME

【出所】NETATMO WELCOME［NETATMO］
https://www.netatmo.com/ja-JP/product/security/welcome

聴覚情報を代替するAmazon EchoとGoogle Home

【出所】Amazon Echo［Amazon.com］
https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/b?node=9818047011

【出所】Google Home［Google］
https://store.google.com/jp/product/google_home

□ カメラやマイクは古くからあるセンサーですが、人工知能と組み合わせることで新たな用途が広がっています。

五感で感知不可能な情報のセンシング

◆電波に基づく位置情報など、五感では感知不可能な情報もセンシングから得ることできます。

- 地磁気や電波など、**人間の五感では感知できない情報も、センシング**によって得ることができます。

人間の五感で感知不可能な情報とセンサー

情報	センサー	用途
電波	電波受信機（GPS・モバイル通信用電波）	位置の把握
地磁気	磁気センサー（電子コンパス）	方角の把握
粒子・花粉	パーティクルカウンター（微粒子計）	空気環境の測定
二酸化炭素	CO2センサー	二酸化炭素濃度の測定、農作物の成長促進
放射能	ガイガーカウンター（放射線量計測器）	放射線環境の測定

- 測定した電波の情報を変換することで位置情報を得るなど、収集した情報と用途が異なるケースもあります。
- 電波による位置の把握は、GPSや携帯電話の基地局との通信状況に基づく位置情報の把握に加えて、Bluetoothをはじめとするビーコンによるものがあります。
  - GPS（Global Positioning System）は衛星測位システムを用いて、衛星から受信した電波に基づく位置情報の把握を指しています。
  - ビーコンとは狼煙（のろし）を原義とし、無線通信においては「電波を利用として、位置情報などを取得する仕組み」を指しており、通信の有無や信号強度に基づいて位置を特定します。
- Bluetoothは通信距離が最大数十メートルであるため、GPSや携帯電話の基地局との通信状況に基づく位置情報に比べて高精度の把握が可能です。
  - 本講座のリーディング技術において紹介したRFIDも、高精度の位置情報の把握に利用することができます。

□ 通信事業者による位置情報の把握と活用は、講座1-3においても説明します。

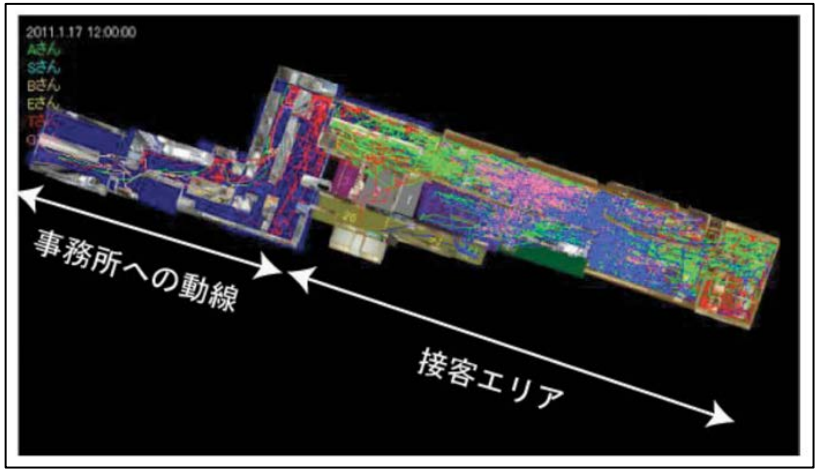
高精度の位置情報のセンシング事例（店舗内の位置把握）

◆センシング技術を活用することで屋内における詳細な位置・移動情報を把握できます。

- がんこフードサービスは、（独）産業技術総合研究所と共同で外食サービスにおけるデータ収集、分析を行いました。
- 分析の一環として、飲食店の従業員にセンサーを取り付け、店舗内の位置、移動を可視化しました。
  - 磁気センサーで方向、加速度センサーで移動の動きを捕捉し、アクティブ型RFIDで位置情報を補正することで移動情報を収集しました。
- 収集したデータのシミュレーションに基づいて改善策を実施することで、飲食店の経営改善をもたらしました。
  - 月商3000万円の店舗で、月間経常利益が300万円の増加があったことが示されています。

従業員の衣服へのセンサーの取り付け

センサーのデータに基づく移動経路の可視化

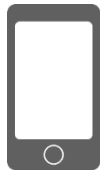


【出所】産総研LINK2017年4月［（独）産業技術総合研究所］
http://www.aist.go.jp/Portals/0/resource_images/aist_j/aistinfo/aist_link/no_12/no_12_full.pdf

□ センサーによって移動や位置情報も正確に測り、可視化することでビジネスでの利活用につながりました。

スマートフォンが持つセンシング機能

◆スマートフォンにおいては、様々なセンシング機能が搭載されています。



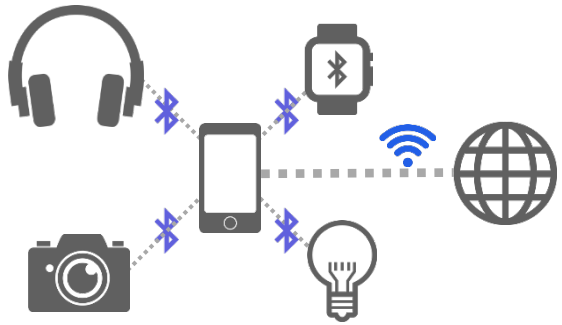
- 普及している**スマートフォン**には、カメラやマイクをはじめとして、**様々なセンサーが搭載**されています。

スマートフォンに搭載されているセンサー

情報	センサー	用途
光、画像	照度センサー、カメラ	画面の明るさ調節、撮影用カメラ
音声	マイクロフォン	通話、音声入力
圧力	圧力センサー	物理ボタン
生体電流	静電容量式タッチセンサ	タッチパネル、指紋認証
地磁気	磁気センサー（電子コンパス）	地図表示等における方角の把握
電波	電波センサー（GPS電波等の受信）	地図表示等における位置の把握
速度変化（重力加速度）	加速度センサー	機器の向きによる画面方向の変更
角加速度（回転）	ジャイロセンサー	カメラの手ぶれ補正、AR（拡張現実）の表示補正

- スマートフォンは様々なセンシング機能を持ち、インターネットにつながる端末として広く一般に普及していることで、様々な活用が期待されています。

- スマートフォンが仲介役となり、Bluetooth接続などのIoT機器が収集した情報をインターネットに送受信するケースもあります。



□ スマートフォンは「持ち歩くIoT」ですが、続いて「身につけるIoT」としてのウェアラブルデバイスを紹介します。

身につけるIoT(ウェアラブルデバイス)

◆身につけるIoTはウェアラブルデバイスと呼ばれ、腕時計型やメガネ型が代表的です。

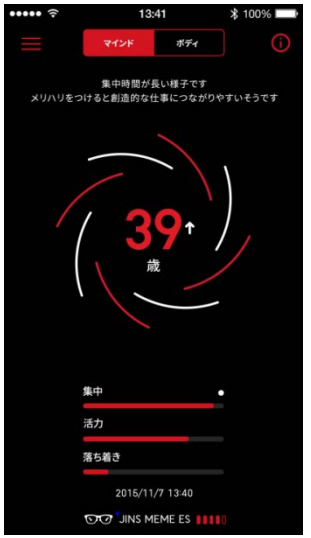
- 腕時計型やメガネ型など、**身につけるIoT機器をウェアラブルデバイス**と呼びます。

- ウェアラブルデバイスは、手に持つ必要がないセンシング機器やコントローラーの役割を果たします。
  - 腕時計型のウェアラブルデバイスの「Apple Watch」は、利用者の心拍数のセンシングができ、音楽再生などのコントローラーとしても利用できます。
    - Apple Watchは、血管における緑色光の吸収量を光センサーでセンシングすることによって脈拍を測定しています。
  - メガネ型のウェアラブルデバイスの「JINS MEME（ジンズ・ミーム）」は、利用者の眼球の動きや体の動きのセンシングができます。
    - JINS MEMEでは、センシングで得られた情報から、集中力や眠気を測定することができます。

心拍数の計測ができるApple Watch
（腕時計型ウェアラブルデバイス）

【出所】Apple Watch［Apple Inc.］
https://support.apple.com/ja-jp/HT204666

眼球・体の動きが感知できるJINS MEME
（メガネ型ウェアラブルデバイス）



【出所】JINS MEME ES［株式会社ジンズ］
https://jins-meme.com/ja/products/es/

ウェアラブルデバイスにおけるセンシング

◆心身に関する情報の感知など、ウェアラブルデバイスには特徴的なセンシングがあります。

- ウェアラブルデバイスの用途は**「心身に関する情報収集」「位置や速度に関する情報収集」「入力・運動支援」**に大別できます。

ウェアラブルデバイスにおけるセンシング例

情報	主なセンサー	用途の大分類	代表的な用途
血流・心拍数	光センサー	**心身に関する情報収集**	健康管理、ストレス、病気の予兆検知
体温	赤外線センサー		健康管理、病気の予兆検知
脳波	頭部電位センサー		ストレス、睡眠深度の測定
眼球の運動	アイトラッカー		集中力の測定、注目対象の把握
位置情報	電波センサー（電波発信もあり得る）	**位置や速度に関する情報収集**	位置の測定
移動情報	加速度・ジャイロセンサー		速度の測定
体の動き	加速度・圧力センサー	**入力・運動支援**	身振りやクリックによる操作
筋肉の動き	生体電位センサー		動作検知、ロボットスーツ

- センシング以外のウェアラブルデバイスの機能として、腕時計型のデバイスが振動して情報を通知するケース、メガネ型のデバイス上に情報を表示するケースもあります。
  - ウェアラブルデバイスが得た情報を人間に感知（センシング）させるための作動や表示をしています。

□ メガネ型のウェアラブルデバイス上の情報表示は、講座1-3のMR（Mixed Reality：複合現実）、VR（Virtual Reality：仮想現実）においても、紹介します。

ウェアラブルデバイスによる「心身に関する情報収集」

◆ウェアラブルデバイスの用途として「心身に関する情報収集」が挙げられます。

- **「心身に関する情報収集」**を行うウェアラブルデバイスでは、睡眠時の動きや呼吸数の情報等を収集できます。

- Fitbit社が発売しているリストバンド型ウェアラブルデバイスのAlta HR（アルタ・エイチアール）では、心拍センサー・加速度センサーによって睡眠の深さ、睡眠時間をはじめとした情報を収集できます。

- ミツフジ株式会社が開発している衣服型ウェアラブルデバイスのhamon（ハモン）は、導電性の高い銀繊維を利用して、心電、筋電、呼吸数などの情報を収集できます。

リストバンド型ウェアラブルデバイスAlta HR

【出所】Alta HR［Fitbit］
https://www.fitbit.com/jp/altahr

衣服型ウェアラブルデバイスのhamon

【出所】hamon［ミツフジ株式会社］
https://www.mitsufuji.co.jp/product/hamon/

ウェアラブルデバイスによる「位置や速度に関する情報収集」

◆ウェアラブルデバイスの用途として「位置や速度に関する情報収集」が挙げられます。

- **「位置や速度に関する情報収集」**を行うウェアラブルデバイスでは、利用者自身やスポーツ選手の位置や移動速度を計測することができます。
  - セイコーエプソン株式会社が発売している「Wristable GPS for Trek」は、GPS、地磁気、加速度、温度、気圧のセンサーを備えており、登山において、緯度経度・高度の情報を把握できます。
  - アメリカのプロフットボールリーグのNFLでは、選手の肩パッドにRFIDを埋め込み、選手の位置や速度をリアルタイムに収集し、視聴者の画面上に表示することができます。
    - 下側の画像では、タッチダウンをした選手のトップスピードが時速21.19マイル（→時速34.1km）であったことを示しています。

Wristable GPS for Trek

【出所】セイコーエプソン株式会社
https://www.epson.jp/products/wgps/trek/

ウェアラブルデバイスによる走行速度の把握

【出所】National Football League
http://www.nfl.com/videos/next-gen-stats/0ap3000000895165

ウェアラブルデバイスによる「入力・運動支援」

◆ウェアラブルデバイスの用途として「入力・運動支援」が挙げられます。

- 「**入力・運動支援**」を行うウェアラブルデバイスとして、指輪型のコントローラーやパワードスーツが挙げられます。

- コクヨ株式会社が発売しているフィンガープレゼンター黒曜石は、指にはめてプレゼンテーションを操作できる指輪型のコントローラーです。

- サイバーダイン株式会社のロボットスーツのHAL（Hybrid Assistive Limb：ハル）は、運動を支援するウェアラブルデバイスであり、生体電位を感知し、筋組織の動きを予測することで人間の動作を支援します。

フィンガープレゼンター黒曜石

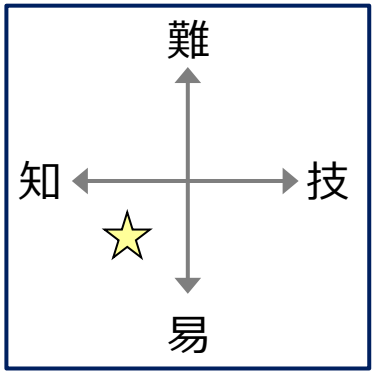
【出所】フィンガープレゼンター黒曜石（コクヨ株式会社）
http://www.kokuyo-st.co.jp/stationery/fp/

生体電位を読み取って運動支援を行うHAL

【出所】HAL（CYBERDYNE株式会社）
https://www.cyberdyne.jp/products/HAL/index.html

□ 運動支援を行うロボットスーツは、ロボットに分類されることもあり、講座1-4でも紹介します。

総務省　ICTスキル総合習得教材



［コース1］データ収集

1-3：位置情報の活用とxR

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_1_3.pdf

	1	2	3	4	5
［コース1］データ収集			●		
［コース2］データ蓄積					
［コース3］データ分析					
［コース4］データ利活用					

本講座の学習内容［1-3：位置情報の活用とxR］

【講座概要】

- スマートフォンの普及状況と通信事業者による位置情報の活用事例を紹介します。
- ウェブサイトやスマートフォンアプリにおける位置情報の活用事例を紹介します。
- 製品化がされているxRを構成するAR、MR、VRのそれぞれの特徴を説明します。
- AR、MR、VRのそれぞれの利用事例を紹介します。

【講座構成】

座学

- [1] 位置情報の把握と活用
- [2] xR(AR・MR・VR)
- [3] AR（拡張現実）
- [4] MR（複合現実）
- [5] VR（仮想現実）

【学習のゴール】

- ✓ スマートフォンの普及状況と通信事業者による位置情報の活用事例を把握する
- ✓ AR（拡張現実）、MR（複合現実）、VR（仮想現実）の特徴を紹介できる。
- ✓ AR、MR、VRのそれぞれの製品例、実用例を理解する。

IoTのThings（モノ）としてのPC、スマートフォン

◆PCやスマートフォンは、一般世帯に最も普及しているIoTです。

- PCやスマートフォンではインターネットにつながっていることで、ウェブサイトの閲覧やメールの送受信ができます。
  - PCはPersonal Computer（パーソナル コンピューター）の略で、日本語では「パソコン」とも略されます。
- PCやスマートフォンには、カメラやマイクをはじめとする様々なセンサーが付いており、センサーとインターネット接続を組み合わせた活用を行うこともできます。
- スマートフォンは、持ち歩くことが前提となっているため、位置情報の把握にも利用することができます。
  - スマートフォンを通信可能な状態にすると、画面にスマートフォンの位置情報を表示したり、家族などの認証を受けた外部ユーザーそのスマートフォンの位置を把握することが可能です。

PC（パーソナル コンピューター）

スマートフォン

家族との位置情報の共有

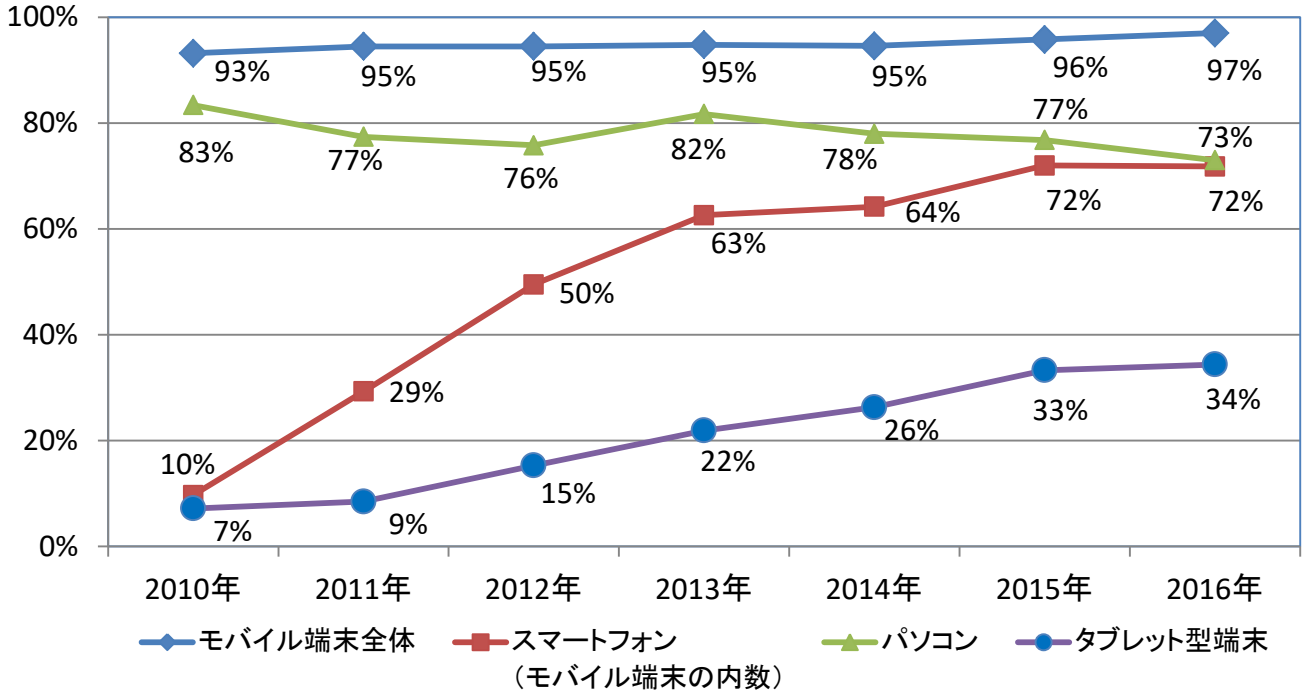
【出所】家族と位置情報を共有する［Apple Inc.］
https://support.apple.com/ja-jp/HT201087

通信機器の世帯保有率

◆パソコン（PC）およびスマートフォンの世帯保有率は2016年において73%となっています。

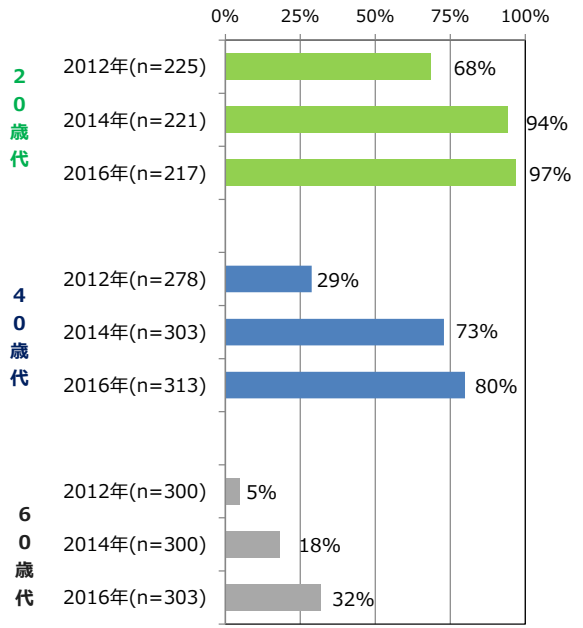
- 総務省が実施している通信利用動向調査によると、2016年におけるPC（パソコン）の世帯保有率は73.0%、スマートフォンの世帯保有率は71.8%が、一般世帯に普及しているIoT機器です。
- パソコンの世帯保有率が減少傾向であるのに対して、スマートフォンの世帯保有率は増加傾向にあります。
- 個人でのスマートフォンの利用率は年齢層によって異なりますが、いずれの年齢層でも増加しています。

通信機器の世帯保有率



【出所】通信利用動向調査［総務省］より作成
http://www.soumu.go.jp/johotsusintokei/statistics/statistics05a.html

年齢層別のスマートフォンの個人利用率



【出所】情報通信メディアの利用時間と情報行動に関する調査報告書［総務省 情報通信政策研究所］より作成
http://www.soumu.go.jp/menu_news/s-news/01iicp01_02000064.html

モバイル通信事業者による位置情報の把握

◆モバイル通信事業者は、モバイル端末の位置情報から人の集まりや移動を把握できます。

- スマートフォンをはじめとするモバイル端末は、電波の送受信が可能な状態にすると、モバイル通信事業者はその端末の位置を概ね把握することができます。
- モバイル通信事業者は、GPS（Global Positioning System）および携帯電話の基地局、Wi-Fiアクセスポイントとの通信状況といった複数の情報に基づいて、モバイル端末の位置を特定します。
- GPSとも呼ばれる人工衛星によるモバイル端末の位置特定は、3つの人工衛星からの電波をモバイル端末が受信し、モバイル端末が地表上にあると仮定することで可能となります。

通信事業者による位置情報把握の方法

【出所】位置情報プライバシーレポート（P67）［総務省］
http://www.soumu.go.jp/menu_news/s-news/01kiban08_02000144.html

GPS（人口衛星によるモバイル端末の位置特定）

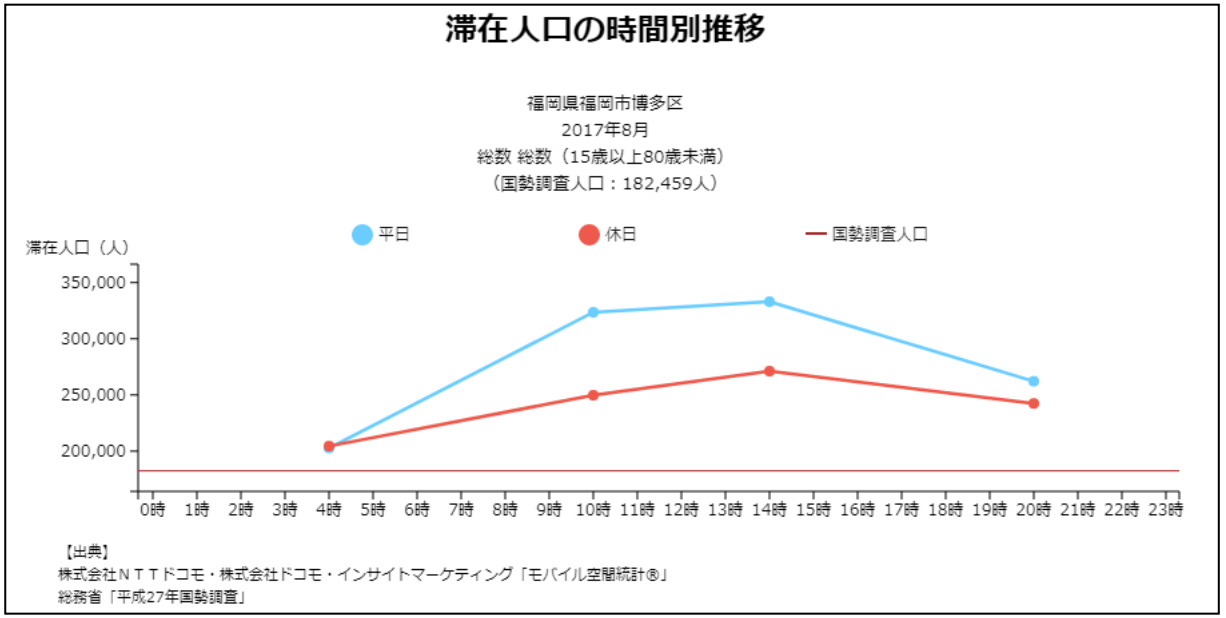
1つめの人工衛星からの電波
⇒ 発信した衛星を中心とする球面でのモバイル端末の位置の絞り込み
2つめの人工衛星からの電波
⇒ 2つの球面が重なる円周でのモバイル端末の位置の絞り込み
3つめの人工衛星からの電波
⇒ 円周と3つめの球が重なる2点でのモバイル端末の位置の絞り込み
モバイル端末が地表上に存在する仮定（または4つめの電波）
⇒ モバイル端末の1点での位置特定

RESASによる位置情報の可視化と活用

◆RESASを利用すると、モバイル通信事業者等が収集した位置情報を可視化できます。

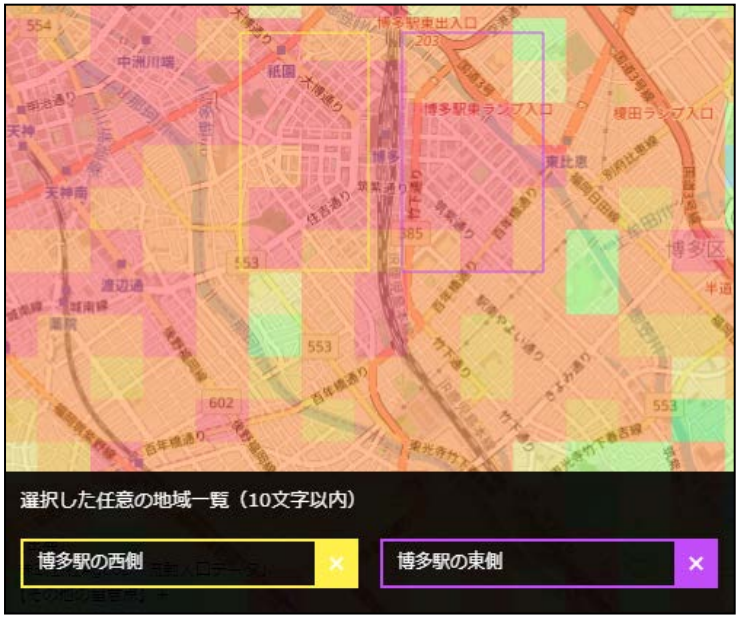
- 内閣官房（まち・ひと・しごと創生本部事務局）が運営するRESAS（リーサス：地域経済分析システム）を利用すると、モバイル端末の位置情報の集計値を無料で可視化することができます。
  - RESASに関する説明および操作実習は、講座4-2にて行います。
- RESASのまちづくりマップの「滞在人口率」から、市区町村単位で時間帯別に人の集積状況を確認したり、「流動人口メッシュ」から、メッシュ単位（駅の東西など）での人の集積状況の違いを把握することができます。
  - メッシュ（mesh）は「網の目」を意味する英単語で、右下の画像のように地図上に網の目状に人の集積状況の濃淡を表すことができます。
  - 市区町村単位、メッシュ単位で人の集積状況を把握することで、出店計画などのビジネスや地域振興の取り組みに活用することができます。

RESASにおける滞在人口率の表示（福岡県福岡市 博多区）



【出所】RESAS（まちづくりマップ：滞在人口率）https://resas.go.jp/

流動人口メッシュの表示（博多駅の東西に着目）



【出所】RESAS（まちづくりマップ：流動人口メッシュ）

ウェブサイト、アプリ管理者による位置情報の活用

◆ウェブサイトやスマートフォンアプリ管理者も端末の詳細な位置情報を活用できます。

- ウェブサイトやスマートフォンアプリの管理者は、閲覧者・利用者の位置情報を把握、活用することができます。
  - ウェブサイトやスマートフォンアプリが詳細な位置情報をユーザーに要求する場合は、「位置情報を管理者に通知して良いですか？」といった旨の確認がウェブブラウザやアプリに表示されます。

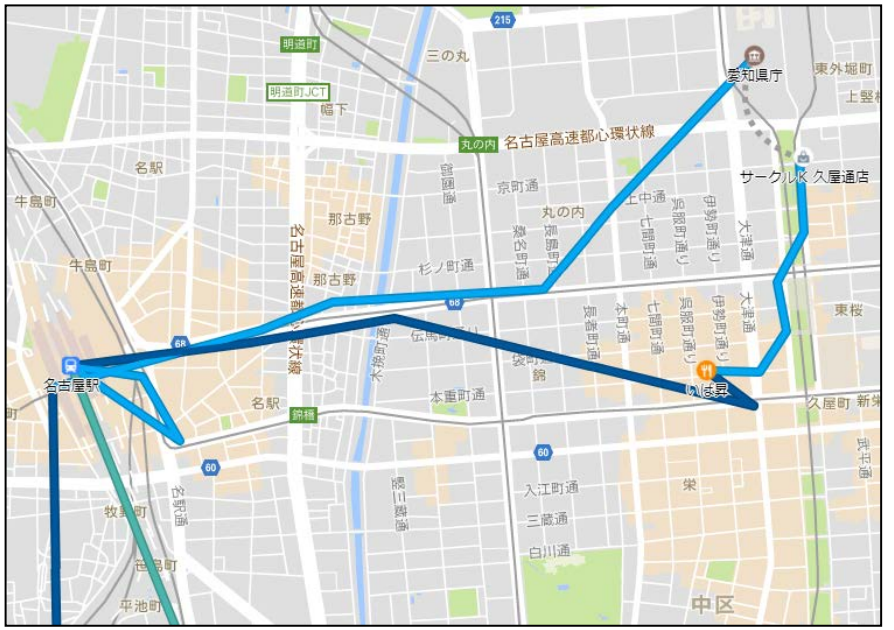
- **ジオコーディング**と呼ばれる変換によって、住所・建物名と位置情報の一方から他方を導出することができます。

  ［変換例：東京都千代田区霞が関2−1−2 中央合同庁舎第2号館
  ⇔緯度35.675514、経度139.75143］

- Googleマップのタイムラインでは、各時点の位置情報から訪問箇所を推定し、移動経路を表示することができます。

Googleマップによる位置情報の地図表示

Googleマップのタイムラインによる移動経路の表示



© 2017 Google Inc, used with permission

リアルタイムの位置情報の活用事例

◆モバイル端末のリアルタイムの位置情報を活用しているスマートフォンアプリもあります。

- モバイル端末のリアルタイムの位置情報の活用例として、「**モバイル端末を持つ利用者自身の位置情報**」および「**他の移動体の位置情報**」のそれぞれの活用が挙げられます。

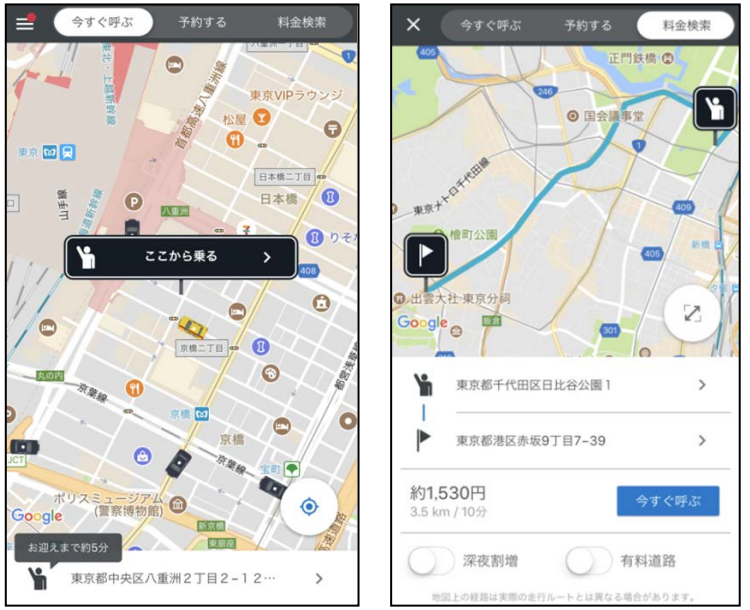
- 「**利用者自身の位置情報**」のリアルタイムの活用例として、スマートフォンアプリ「食べログ」におけるユーザーの現在位置周辺のレストランの検索が挙げられます。

ユーザー位置の周辺情報の検索
(「食べログ」の現在位置周辺のレストラン検索)

【出所】食べログ［株式会社カカクコム］

- 「**利用者自身の位置情報**」および「**他の移動体の位置情報**」のリアルタイムの活用例として、スマートフォンアプリ「全国タクシー」でのユーザー位置とタクシー位置の表示が挙げられます。

位置の通知によるタクシーの呼び出し
(全国タクシーでのユーザー位置とタクシー位置の表示)



【出所】全国タクシー［JapanTaxi株式会社］

位置情報を利用したスマートフォン用ゲーム

◆ゲームアプリ「ポケモンGO」は、プレイヤーの現実の位置情報等をゲームに反映しています。

- 2016年に発売されたスマートフォン用ゲームアプリの「ポケモンGO」は、現実世界におけるプレイヤーの位置情報をゲームにも反映しています。
  - 現実世界にある建物の近くでゲームを行えば、ゲーム内でもその建物が表示されるケースがあります。

ポケモンGOにおける位置情報の活用

©2018 Niantic, Inc. ©2018 Pokémon. ©1995-2018 Nintendo

- 各位置における天気情報を反映することで、現実の場所における天気でゲーム内のキャラクターの出現傾向が変化します。
  - 雪が降った地域には、雪に関連するキャラクターが出現するようになります。

ポケモンGOにおける位置ごとの天候情報の活用

【出所】Pokémon GO Japan
https://www.youtube.com/watch?v=RCB66gZJGqw

□ 「ポケモンGO」では、続くパートで紹介するxRに含まれるAR（拡張現実）の技術も活用しています。

xR（エックスアール）

◆xRは、総じて現実には存在しないものを表現、体験できる技術です。

- 現実世界において実際には存在しないものを、表現・体験できる技術を総称して**xR（エックスアール）**といいます。
  - ICT関連用語では、様々な文字が入る部分に「x」を入れることで、総称として示す場合があります。他の事例として、講座1-1に示したX-tech(エックステック)、講座2-2に示すXaaS（ザース）が挙げられます。

- 実用化済のxRは、**AR（拡張現実）**、**MR（複合現実）**、**VR（仮想現実）**に大別できます。
  - 前スライドに示した「ポケモンGO」において活用されているAR（拡張現実）は、xRと総称される技術の一つです。
  - 実用化された時間的な順番はVR、AR、MRですが、本教材では没入感とも呼ばれる「対象に夢中になる程度」が浅い順からAR、MR、VRの順に紹介します。

xRに関する包含関係

総称としてのxR（○○○ Reality）

総じて、現実世界には存在しないものを表現、体験できる技術

拡張現実 AR（エーアール）	複合現実 MR（エムアール）	仮想現実 VR（ブイアール）
（Augmented Reality）	（Mixed Reality）	（Virtual Reality）
現実世界（一部）に仮想の情報を重ね合わせる。	現実世界（視界全面）に仮想の情報を重ね合わせる。	現実世界の情報は遮断して、仮想世界のみを描く。

AR、MR、VRの特徴

◆AR、MR、VRでは、主な利用機器、一般的な操作方法、特徴が異なっています。

	拡張現実　AR （Augmented Reality）	複合現実　MR （Mixed Reality）	仮想現実　VR （Virtual Reality）
表現	現実世界（一部）に仮想の情報を重ね合わせる。	現実世界（視界全面）に仮想の情報を重ね合わせる。	現実世界の情報は遮断して、仮想世界のみを描く。
主な機器	・スマートフォン ・タブレット型端末 	・メガネ（グラス） ・ヘッドマウントディスプレイ 	・ヘッドマウントディスプレイ
一般的な操作方法	◆ タッチパネル操作や機器を動かす。 	◆ 指などのジェスチャーや空間に現れたパネルを触る。 	◆ 物理的なコントローラーを振ったり、ボタンを押す。
特徴・用途	◆ 記号等を動作開始・位置合わせの鍵（マーカー）として、画像表示等ができる。 ◆ 長さなどの現実世界に関する情報を表示できる。	◆ 複数人で情報を共有でき、説明や協力作業に向いている。 ◆ 指のジェスチャー操作、空間上のパネル操作ができる。	◆ 現実以上の美しさや空想の仮想世界も描画できる。 ◆ 災害発生や遠方への旅行など現実に発生しにくいイベントを体験できる。

□ MRやVRはメガネやヘッドマウントディスプレイを身につけるため、講座1-2で示したウェアラブルデバイスでもあります。

拡張現実　AR（Augmented Reality）

◆ARでは、スマートフォン等を利用して現実の映像の手前にコンピューター画像を表示します。

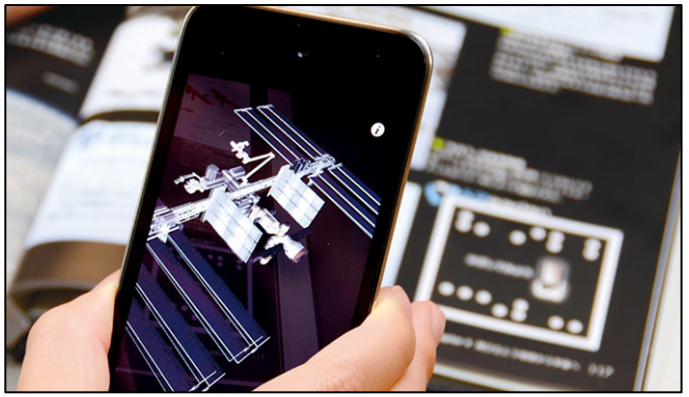
- **拡張現実AR**は、スマートフォンやタブレット型端末を利用して、カメラ等から入力された実際の映像の手前にコンピューター画像を表示する技術です。

- スマートフォンアプリの「AReader（エアリーダー）」では、動作のきっかけとなる専用マーカーをカメラで読み込むことで、現実の映像の手前に3D画像や動画を表示します。

- スマートフォンアプリの「HP Reveal」では、利用者が撮影した写真をマーカーとするARを設定ができます。
  - 下の画像では「総務省 ICTスキル総合習得プログラム」の画像をマーカーとしています。

- スマートフォンアプリの「AirMeasure」では、現実の映像の手前にARの定規を表示する形で、現実の物体の長さを測定することができます。

AReaderによる宇宙ステーションの表示

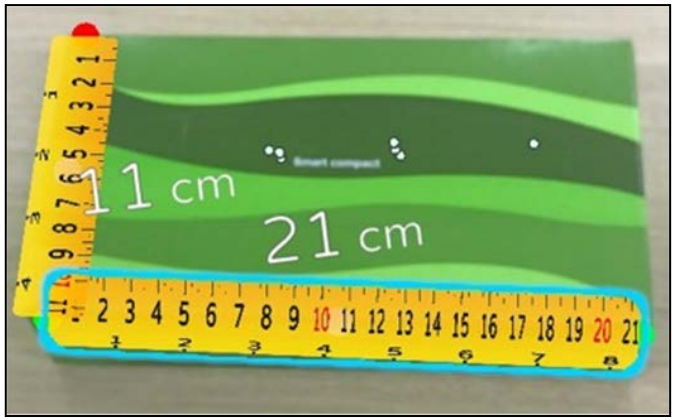


【出所】AReader［凸版印刷株式会社］
https://www.toppan.co.jp/solution/service/AReader.html

HP RevealによるAR表示

【出所】HP Reveal［Aurasma］
https://www.hpreveal.com/

AirMeasureによる箱のサイズのAR測定



【出所】AirMeasure［Laan Labs］
http://armeasure.com/

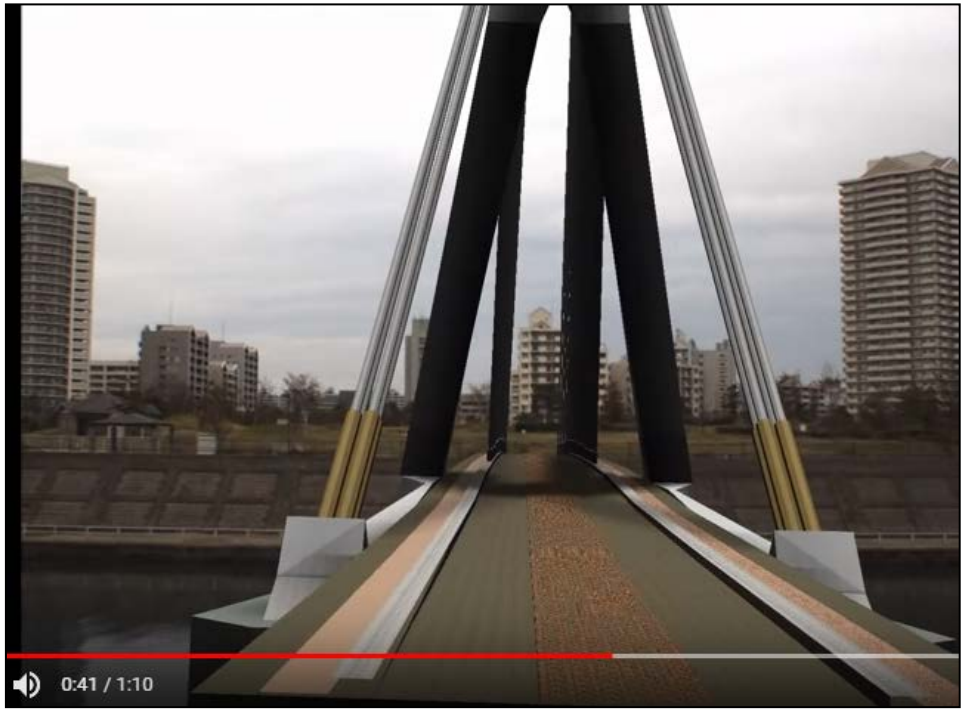
□ ARは、現実の映像を背景として、3D画像・動画、長さなどの現実世界に関する情報を表示することができます。

ARの実用利用

◆ARは、景観シミュレーションや動く写真として実用的に利用されています。

- ARを使うことで、建築予定の橋梁の建築後の景観をシミュレーション映像で示すことができます。
  - 表示する橋梁を3D画像とすることで、合成写真とは異なり、様々な角度からの景観を確認することができます。

ARによる景観シミュレーション



【出所】千葉大学大学院工学研究科 平沢研究室
https://www.youtube.com/watch?v=mkRupfdm22M

- モバイルプリンタのLifeprint（ライフプリント）では、写真をマーカーとし、アプリによって動画を表示することで、スマホで見ると動く写真を印刷することができます。

ARによる動く写真

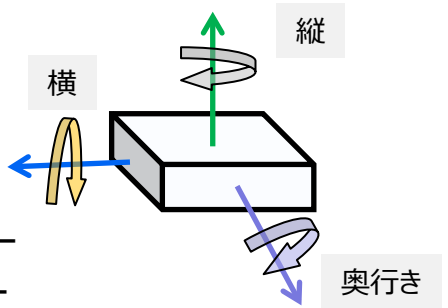
【出所】Life Print
https://www.youtube.com/watch?v=aRsmHpDSSQo

ARのゲーム・娯楽利用

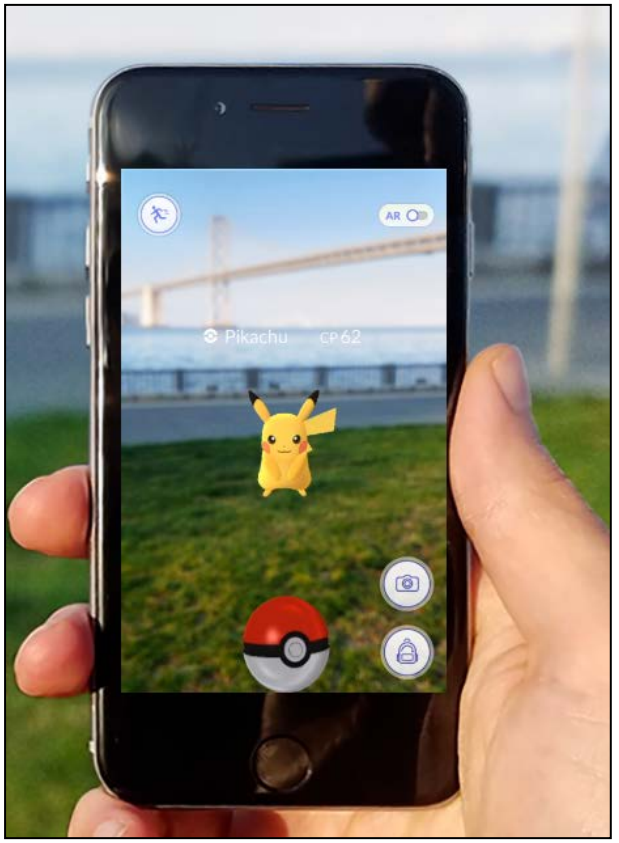
◆「ポケモンGO」では、ポケットモンスターをAR（拡張現実）によって表示することができます。

- スマートフォン向けゲーム「ポケモンGO」は、ジャイロセンサーを搭載したスマートフォンによって、現実の背景の手前にARでポケットモンスターを表示可能です。
  - ジャイロセンサー非搭載のスマートフォンで、ポケモンGOをプレイすると、AR表示ができずに、コンピューターが画像の背景になります。
- スマートフォンには動きを把握する**加速度センサー**に加えて、角加速度（回転）を把握する**ジャイロセンサー**が搭載されているものがあります。
- 3軸の加速度センサーと3軸のジャイロセンサーを備えて、全方向への動きと回転が把握できるセンサーを**6軸センサー**といいます。
  - 縦、横、奥行の3軸の加速度センサーに加えて、縦回転、横回転、奥行側の回転の3軸のジャイロセンサーが搭載されていれば、全方向への動きと回転が感知できます。
- 6軸センサーによって、スマートフォンの動かして角度を変えても、現実の背景と3D画像を正しく重ね合わせることができ、精度の高いAR表示が可能になります。
- 「ポケモンGO」のARにおける現実の背景とコンピューター画像の自然な合成は、スマートフォンのセンサーの技術に支えられています。

6軸センサーの各軸の感知



ポケモンGOのAR表示



©2017 Niantic, Inc. ©2018 Pokémon. ©1995-2018 Nintendo

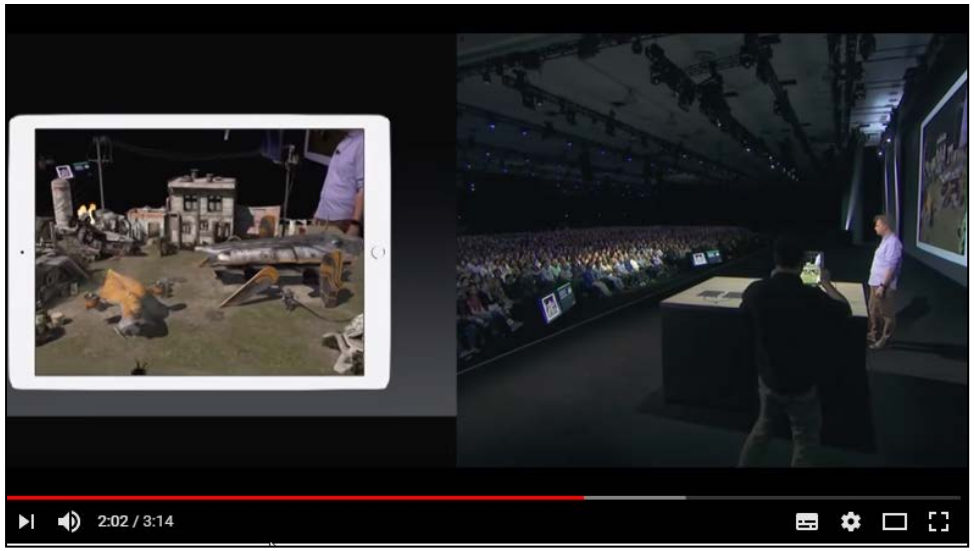
□ スマートフォンはカメラ搭載のみならず、センサーによって動きや回転を感知できることから、AR活用に適した機器です。

ARの開発環境の整備

◆ARアプリを開発しやすい環境、3Dオブジェクトの提供サイトも整えられています。

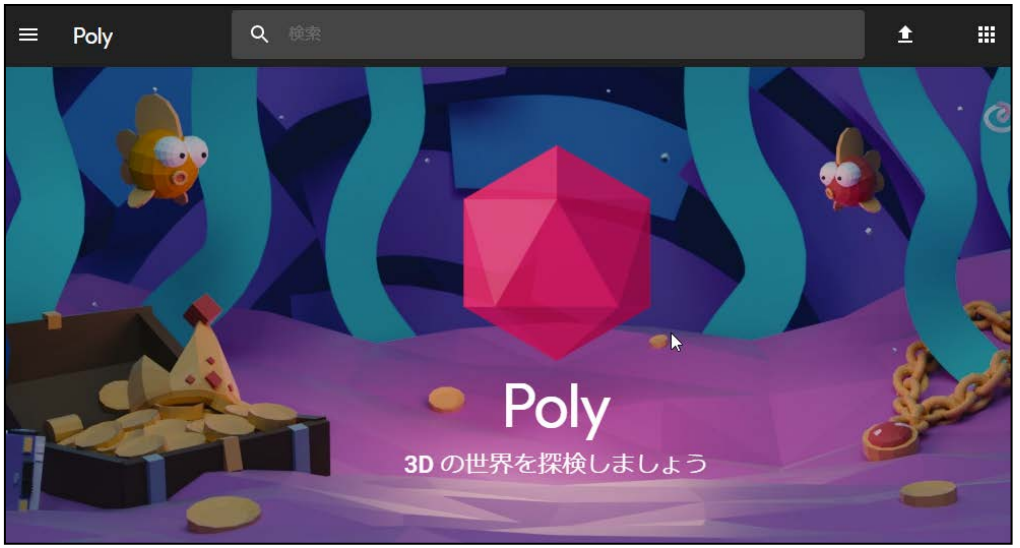
- Appleが2017年6月に開催したWWDC（The Apple Worldwide Developers Conference）2017において iOS11におけるARの開発を容易にする「AR Kit」を発表しました。
- Appleは「iOSは世界最大のARプラットフォームになります」という発表しました。
- Googleは2017年11月1日、ARやVRにおいて無料で使える3Dオブジェクトを集めたカタログサイト「Poly」を立ち上げたこと発表しました。
- 3Dオブジェクトはクリエイティブ・コモンズで公開されており、出所の表記をすれば、誰でも利用することができます。
  - クリエイティブ・コモンズに関しては、講座4-1にて紹介します。

WWDC 2017のAppleによる「AR Kit」紹介



【出所】Unreal Engine
https://www.youtube.com/watch?v=S14AVwaBF-Y

Googleが運営する3Dオブジェクトの提供サイトPoly



【出所】poly［Google］
https://poly.google.com/

複合現実　MR（Mixed Reality）

◆MRは、視界全体にARの表現ができることに加えて、複数名での同時体験が可能です。

- **MR（複合現実）**は、メガネ（グラス）等を通して見る視界全体のARのイメージです。
  - ARのように背景に現実世界を写すことも、VRのように仮想世界のみを写すことも可能です。
- MRはマイクロソフト社が「Windows Mixed Reality」として推進しており、代表的な機器としてHoloLens（ホロレンズ）が挙げられます。
  - Windows MR用のヘッドマウントディスプレイとしては、Immersive（没入的）Deviceと呼ばれるVR用ヘッドマウントディスプレイに近い性質の製品もあります。
- MRでは複数名での同一の映像を確認ができ、説明や協力作業に適しています。

Microsoft HoloLens

MRによる情報共有、空間上のパネル操作のイメージ

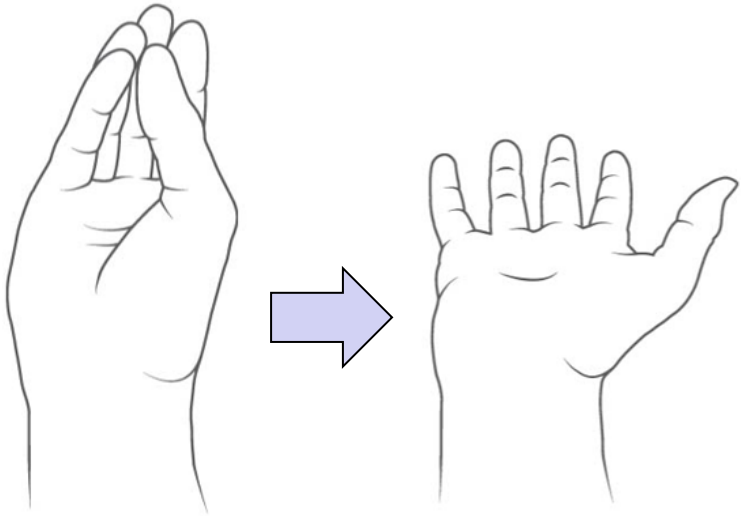
【出所】Microsoft HoloLens［日本マイクロソフト株式会社］
https://www.microsoft.com/ja-jp/hololens

HoloLensにおけるジェスチャー操作

◆HoloLensは、頭や手先の動きによるジェスチャー操作が可能です。

- HoloLensの特徴的な操作方法として、頭や手先を動かすことでのジェスチャー操作が挙げられます。
  - MRでは物理的なコントローラーを振ったり、ボタンを押したりして操作するケースもあります。
- 指先の標準的なジェスチャーとして、上側に手を開く動作の「Bloom（ブルーム）」が設定されています。
  - 一般にメニューを開く動作に割り当てられています。

Bloom（ブルーム）の手の動き



【出所】Use gestures［日本マイクロソフト株式会社］
https://support.microsoft.com/ja-jp/help/12644/hololens-use-gestures

- 指先の標準的なジェスチャーとして、人差し指と親指でつまむ動作の「Tap and Hold(タップアンドホールド)」が挙げられます。
  - マウス操作におけるドラッグに対応しています。

Tap and Hold(タップ アンド ホールド)を行う様子

【出所】Microsoft HoloLens: Gesture Input［Microsoft HoloLens］
https://www.youtube.com/watch?v=kwn9Lh0E_vU

MRの実用利用

◆MRは、コンピュータ表示を同時に見る業務（紹介、協力作業）等に活用されています。

- MRは顧客と同じ画像を見ることができ、紹介作業に適しています。
- 建設現場においては、現実の壁にコンピューター画像による壁内の配水、配電の様子を示すことができます。

MRによって壁内部の配水、配電を紹介する様子

【出所】小柳建設 - Microsoft HoloLens を活用した Holostruction［日本マイクロソフト株式会社］
https://www.youtube.com/watch?v=TmDHZ-NMQe0

- MRでは、体内の骨の状態を表示することができ、整形外科手術のトレーニングに活用することができます。
- 複数名で同じ画像を見ることで、協力作業のトレーニングも可能になります。

MRによって整形外科手術のトレーニングをする様子

【出所】[Microsoft HoloLens] 日本の Mixed Reality パートナーとソリューション［日本マイクロソフト株式会社］
https://www.youtube.com/watch?v=KK42C6OqJvQ

仮想現実　VR（Virtual Reality）

◆VRは、現実世界を遮断して仮想世界を体験する技術で、ゲームなどで利用されています。

- **VR（仮想現実）**は、視界全面を覆うヘッドマウントディスプレイで仮想世界を体験する技術です。
  - ヘッドマウントディスプレイは透過するレンズがあるグラスとは異なり、現実の背景が見えません。
- VRは、ARやMRに比べて古くから製品化され、ゲームなどで利用されています。
  - VRは、現実世界を重ねて表示しないため、基本技術はARやMRよりも相対的に易しくなっています。
  - 1995年に任天堂より発売されたVIRTUAL BOY（バーチャルボーイ）がVRゲーム機の先駆けとなっています。
- 2016年10月にはPlayStation 4で仮想現実VRを楽しめる「PlayStation VR」が発売されました。
  - VRヘッドセットに内蔵されたセンサーが頭部の動きを把握することに加えて、VRヘッドセットの表面に搭載されたLEDライトをカメラがとらえることで、頭部の位置を正確に認識します。
  - 子供発育に関する理由から、PlayStation VRの対象年齢は12歳以上とされています。

VIRTUAL BOY（バーチャルボーイ）

【写真提供】任天堂株式会社

PlayStation VRのヘッドセットとカメラ

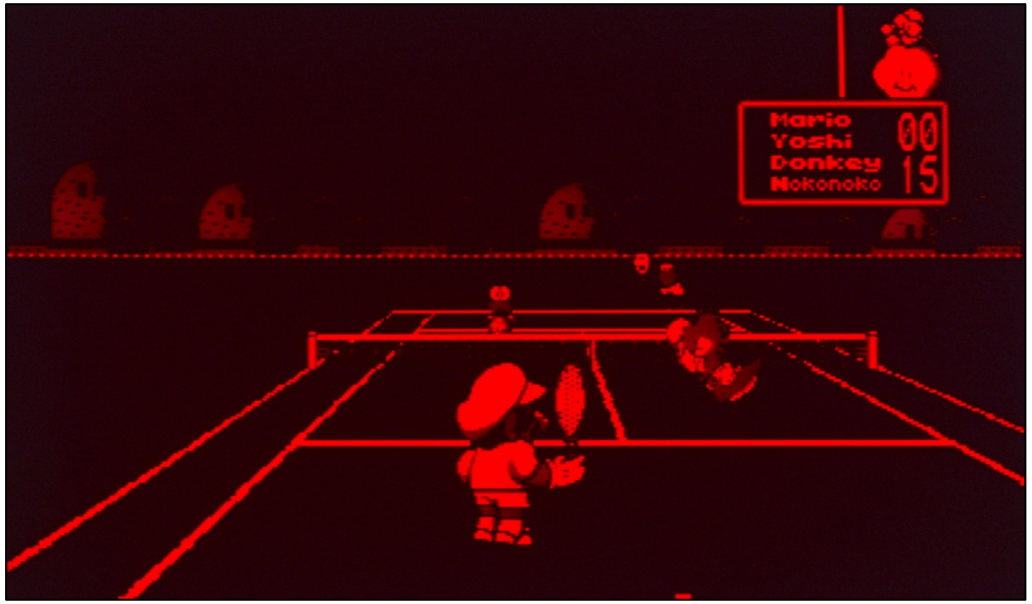
【出所】株式会社ソニー・インタラクティブエンタテインメント
http://www.jp.playstation.com/psvr/product-lineup/playstation-vr-with-pscamera/

VRにおける仮想世界への没入

◆VRを利用すると、現実とは別の仮想世界に没入することができます。

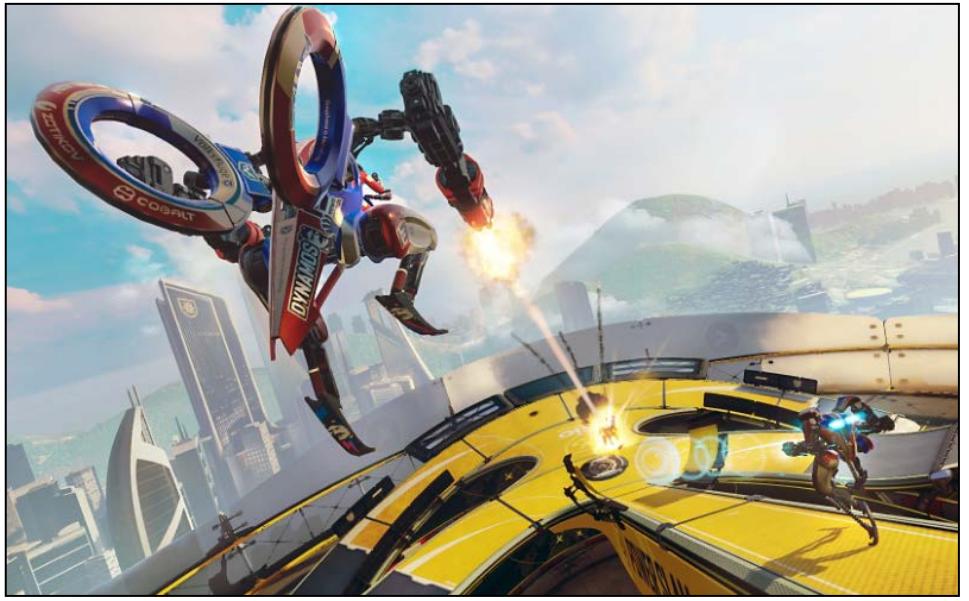
- VRは現実世界の情報を遮断して、仮想世界に入り込むことができます。
  - VRの基本技術は製品化初期の頃から変わりませんが、映像技術の向上から現実と同等の美しい表示が可能となってきました。
- VRはARやMRに比べて、対象に夢中になる**没入感**が高いとされています。
  - デジタル大辞泉（小学館）では、「没入感」の意味として『他のことが気にならなくなるほど、ある対象や状況に意識を集中している感じ。特に、音楽・映画・ゲームのほか、バーチャルリアリティー（VR）などで体験する感覚についていう。』と記しています。
- VRゲームの中にはセンサーを利用して、現実世界の体の動きを仮想世界の動きに反映できるものもあります。

マリオズテニス
（VIRTUAL BOYのゲーム）



(c)1995 Nintendo

RIGS Machine Combat League
（PlayStation VRのゲーム）



【出所】株式会社ソニー・インタラクティブエンタテインメント
https://www.jp.playstation.com/games/rigs-machine-combat-league-ps4/

VRの実用利用

◆VRでは、再現が困難なイベントや、実現が難しい事象の体験にも活用されています。

- VRでは水没車両や津波といった再現の難しいイベントの体験にも利用されています。
  - 下の画像に示した「水没車両からの脱出」の動画は、ヘッドマウントディスプレイを装着する形式のVRではありませんが、リンク先は360度動画として、PCのディスプレイに表示される動画が表示する向きをマウス操作で上下左右に動かすことができます。

水没車両のVR体験（360度動画）

【出所】360度動画でVR体験！水没車両からの脱出【疑似体験版】［日本自動車連盟（JAF）］
https://www.youtube.com/watch?v=CTXGT3Z5s_w

- 2017年10月から2018年2月にかけて国立科学博物館で開催された「古代アンデス文明展」では、「VRウユニ塩湖」として南米ボリビアのウユニ塩湖の体験ができました。
  - VRウユニ塩湖では講座1-5で示す「デバイスWebAPI」技術を使うことで、複数名でのVR動作の共有ができるようになっていました

VRウユニ塩湖

【出所】VRウユニ塩湖［アンデス文明展］
https://www.youtube.com/watch?v=tlhz7TnKTd8

VRのゲーム･娯楽利用

◆VRの利用形態は、スマートフォンを利用する簡便なタイプから専用施設まで様々です。

- ゴーグル型の簡易型VR装置のハコスコでは、視界部分にスマートフォンを入れるだけでVRの体験ができます。
  - 段ボール型のハコスコの価格は1000円台となっており、スマートフォンがあれば、安価にVR体験を行うことができます。
  - ハコスコには推奨年齢を7歳以上とする立体視が可能な二眼モデル、6歳以下の子供でも利用できる一眼モデルがあります。

ハコスコにスマートフォンを入れる様子

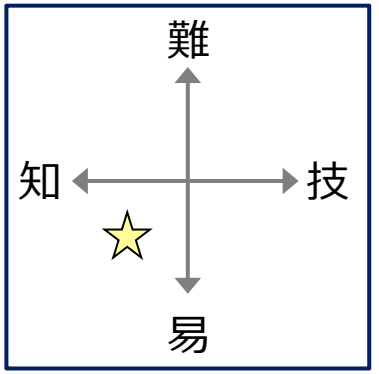
【出所】株式会社ハコスコ
https://hacosco.com/product/

- 2017年7月に開店したエンターテインメント施設「VR ZONE SHINJUKU」では、様々なVRアクティビティを楽しむことができます。
  - 「VR ZONE SHINJUKU」には、7歳以上からプレイできるVRアクティビティ、13歳以上からプレイできるVRアクティビティがあります。

「VR ZONE SHINJUKU」におけるゲームの様子

【出所】株式会社 バンダイナムコ エンターテインメント
https://vrzone-pic.com/

総務省　ICTスキル総合習得教材



［コース1］データ収集

1-4：現実世界へのフィードバックとロボット

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_1_4.pdf

	1	2	3	4	5
［コース1］データ収集				●	
［コース2］データ蓄積					
［コース3］データ分析					
［コース4］データ利活用					

本講座の学習内容［1-4：現実世界へのフィードバックとロボット］

【講座概要】

- データから現実世界へのフィードバックとして、IoT機器に対する通知・制御を説明します。
- フィードバックによって制御されるIoT機器の一種としてのロボットを紹介します。
- 日本における産業用ロボットの活用、市場に関して概説します。
- 屋内用ロボット、屋外用ロボット（フィールドロボット）の利活用を説明します。
- 特に注目を集めているフィールドロボットとして、ドローンに関する利活用や規制を説明します。

【講座構成】

座学

- [1] IoTによる現実世界へのフィードバック
- [2] 産業用ロボット
- [3] 屋内用ロボット
- [4] 屋外用ロボット（フィールドロボット）
- [5] ドローン（多用途の屋外用ロボット）

【学習のゴール】

- ✓ IoT機器のフィードバックの種類と事例を把握する。
- ✓ ロボットの分類としての産業用ロボット、サービスロボットとその用途を理解する。
- ✓ サービスロボット内の分類と利用例を紹介できる。
- ✓ 注目を集めるドローンの普及と規制を把握する。

IoTによる現実世界へのフィードバックの種類

◆IoTによるフィードバックは、通知と制御に分類され、それぞれ身近な例があります。

- IoTによる現実世界へのフィードバックは、**通知**と**制御**に分類されます。
  - 現実世界における作動は「（デジタルからアナログへの）フィードバック」とも言われます。フィードバックは「（次の活用のために）出力を入力側に戻す」ということを意味しています。

フィードバック
- 通知：人間や他の機器に知らせる
- 制御：その機器自体が作動する

通知

制御

- 通知の身近な例は、スマートフォンでメールを受信した際の通知音や振動です。
  - インターネットカメラで遠隔地の画像を送信するケースも、画像情報を通知していると考えることができます。
- 通知を受けて反応するのは人間とは限らず、プログラムや機械であるケースもあります。
  - 自動返信メールは、メールの受信通知やウェブフォームでの入力通知を受けたメールサーバが自動的に返信をしています。
- 複雑な入力に対応する通知としては、入力した文章を解析して反応する**チャットボット**が挙げられます。
- 制御の身近な例としては、外出先からスマートフォンでエアコン等を操作することが挙げられます。
  - 受信した情報によって制御され、動作する機器を作動装置（アクチュエータ）といいます。

複雑な通知ができるチャットボット

外出先から操作できるエアコン

通知を行うIoT事例

◆通知機能のあるIoT機器は、探し物の発見や帰宅・登校の通知などに利用されています。

- ソフトバンクコマース&サービス株式会社が販売している「Tile Mate（タイルメイト）」や「Tile Slim（タイルスリム）」は、スマートフォンから鳴らしたり、スマートフォンを鳴らすことができ、探し物の発見に役立つIoTです。

探し物の発見用IoT（Tile Mate）

【出所】Tile Mate［ソフトバンクコマース&サービス株式会社］
https://www.softbankselection.jp/special/tile/

- Qrio株式会社が販売している「Qrio（キュリオ） ただいまキット」は、子供の帰宅・登校を保護者のLINEアカウントへメッセージで通知することができるIoTです。
  - LINE（ライン）は、テキストチャットを中心とするソーシャル・ネットワーキング・サービスのアプリケーションであり、PCやスマートフォンから利用することができます。

子供の帰宅・登校通知IoT（Qrio ただいまキット）

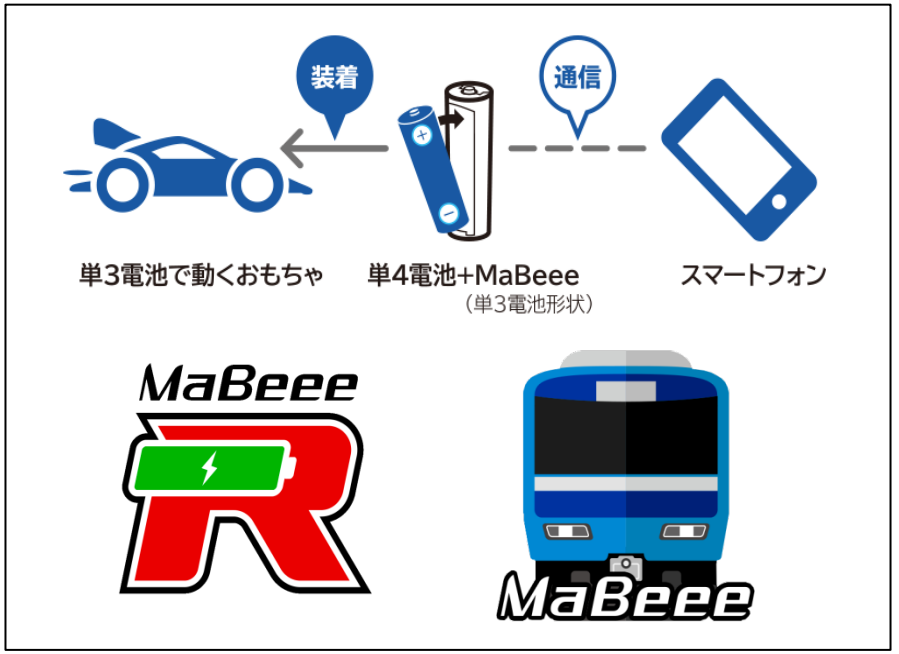
【出所】Qrio ただいまキット［Qrio株式会社］
https://qrio.me/tadaima-kit/

制御を行うIoT事例

◆制御機能のあるIoT機器として、電流制御やシャッターの自動開閉の例が挙げられます。

- 乾電池型のIoT機器「MaBeee(マビー)」は、スマートフォンで電流をコントロールでき、おもちゃの操作ができます。
  - MaBeeeは、Bluetoothで乾電池の電力制御をしており、操作には専用のスマートフォンアプリを利用します。

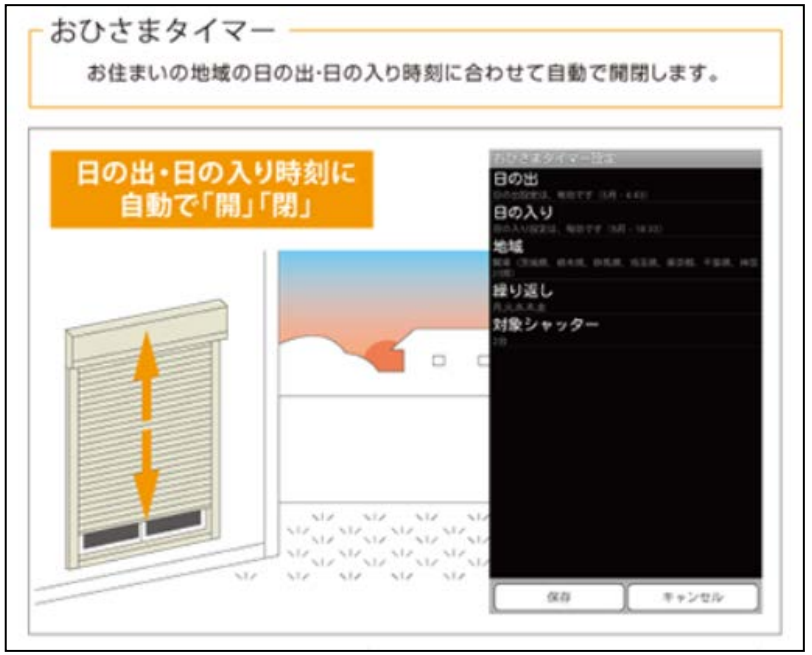
乾電池の電流制御IoT（MaBeee）



【出所】MaBeee［ノバルス株式会社］
https://mabeee.mobi/

- 文化シャッターの「マドマスター・スマートタイプ」は、地域毎の日の出、日の入り時間に合わせた自動開閉を設定できます。
  - 気象警報の情報をインターネットから受信すると、自動的にシャッターを閉じる設定も可能です。

シャッター制御のIoT（マドマスター・スマートタイプ）



【出所】窓シャッター［文化シャッター］
http://www.bunka-s-pro.jp/w-hutter/feature/smartphone/

IoTとしてのロボット

◆ロボットは外部と通信でき、IoTの一つという考え方も根付いてきました。

- 2006年に経済産業省から公表されたロボット政策研究会の報告書では、ロボットの定義として『**「センサー」「知能・制御系」「駆動系」の３つの要素技術**を有する、知能化した機械システム』としていました。
  - 2006年の時点では、インターネットと接続しているロボットは一般的ではありませんでした。

【出所】ロボット政策研究会報告書(2006年5月）［ロボット政策研究会］
http://warp.da.ndl.go.jp/info:ndljp/pid/3486530/www.meti.go.jp/press/20060516002/20060516002.html

- 2014年度において、内閣総理大臣の下に開催された「ロボット革命実現会議」における議論を経て、とりまとめられた**「ロボット新戦略 ―ビジョン・戦略・アクションプラン―」**では、ロボットの劇的変化として3点を挙げています。

2010年代におけるロボットの3種の劇的変化

- 自ら学習し行動するロボットへと**「自律化」**
- 様々なデータを自ら蓄積・活用する**「情報端末化」**
- 相互に結びつき連携するロボットへと**「ネットワーク化」**

⇒ **ロボットのIoT化**

【出所】ロボット新戦略（2015年1月）［日本経済再生本部］http://www.kantei.go.jp/jp/singi/robot/

- NEDOロボット白書2014では、ロボットの定義が流動的な理由として「時代とともに科学技術はもとより、産業構造、社会制度、文化なども変化し、ロボットの役割や受け取られ方も変遷してきていること」を挙げています。

【出所】NEDOロボット白書2014(2014年7月）［国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO：ネド）］
http://www.nedo.go.jp/library/robot_hakusyo.html

□ 通信技術の発達・普及によって、ロボットは外部と通信でき、IoTの一つという考え方も根付いてきました。

ロボットの大分類

◆ロボットは産業用ロボットとサービスロボットの二種類に大別されます。

- ロボットは、工場内で製造に利用される**産業用ロボット**とそれ以外の**サービスロボット**に大別されます。
- **産業用ロボット**については、日本工業規格（JIS）で「自動制御によるマニピュレーション機能または移動機能をもち、各種の作業をプログラムによって実行できる、産業に使用される機械」と規定されています。
  - 国際標準化機構（ISO）では、産業用ロボットを「自動的なコントロール、再プログラミング、多目的で3軸以上の操作が可能で、固定または移動可能かにかかわらず産業自動化（industrial automation applications）に利用される物」と定義しています。
  - 産業用ロボットは、一般に農業や流通業といった非製造業で利用されるロボットを含みません。国際標準化機構（ISO）では、産業用ロボットをモニターアームの構造によって細分化しています。
- 産業用ロボット以外のロボットを総称して**サービスロボット**と呼び、サービスロボットに含まれるものとして掃除用ロボットから人間と会話できるコミュニケーションロボットまで様々です。
  - 国際標準化機構（ISO）では、サービスロボットを業務向けか個人利用向けかで区分しています。

産業用ロボット

コミュニケーションロボット　KIROBO mini

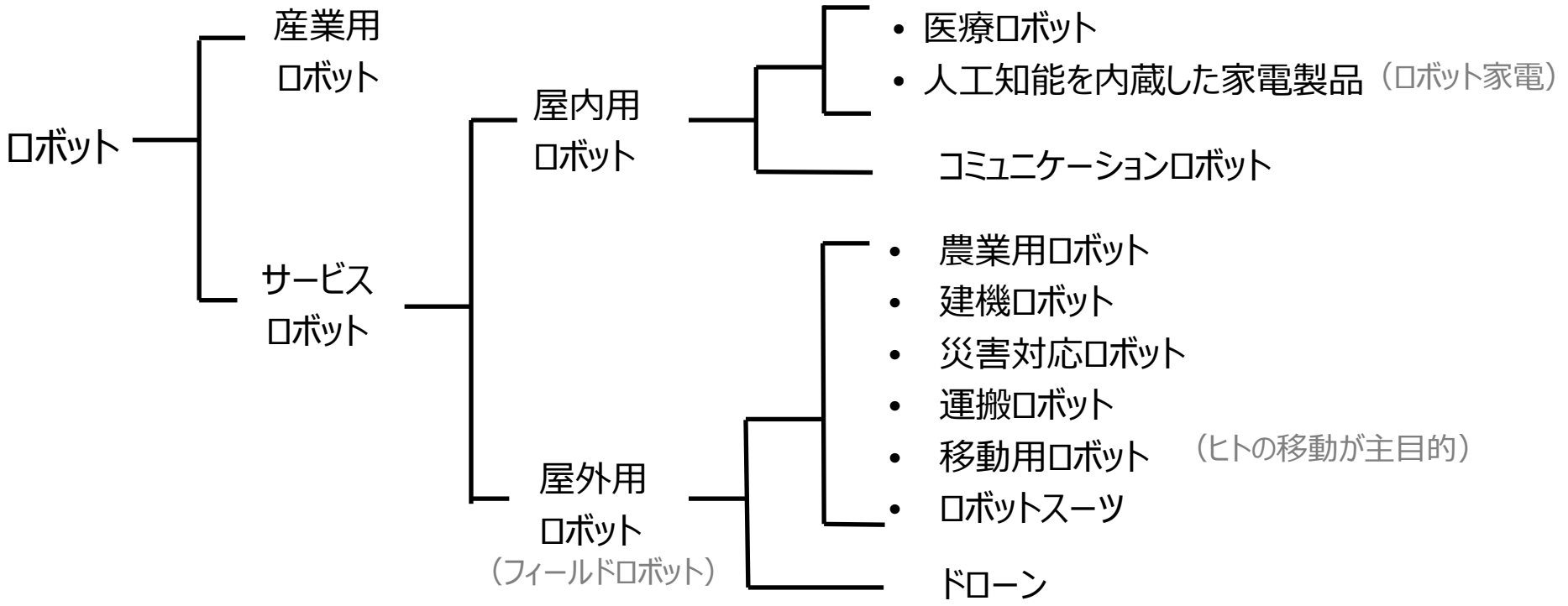
【出所】KIROBO mini［トヨタ自動車株式会社］
https://toyota.jp/kirobo_mini/

ロボットの分類

◆サービスロボットは、主な利用場所に応じて分類することができます。

- サービスロボットは主な利用場所に応じて、下記のように分類することができます。
  - サービスロボット内には、標準化機構等で確立した分類基準がないため、本教材では主たる利用場所に応じて便宜的に分類しています。
  - NEDOロボット白書2014において、「フィールドロボットは、狭義には屋外環境で稼働しているロボット、広義には屋外、もしくは、屋内外で活動する遠隔操作機械として定義される。」と記されています。

ロボットの分類に関する樹形図

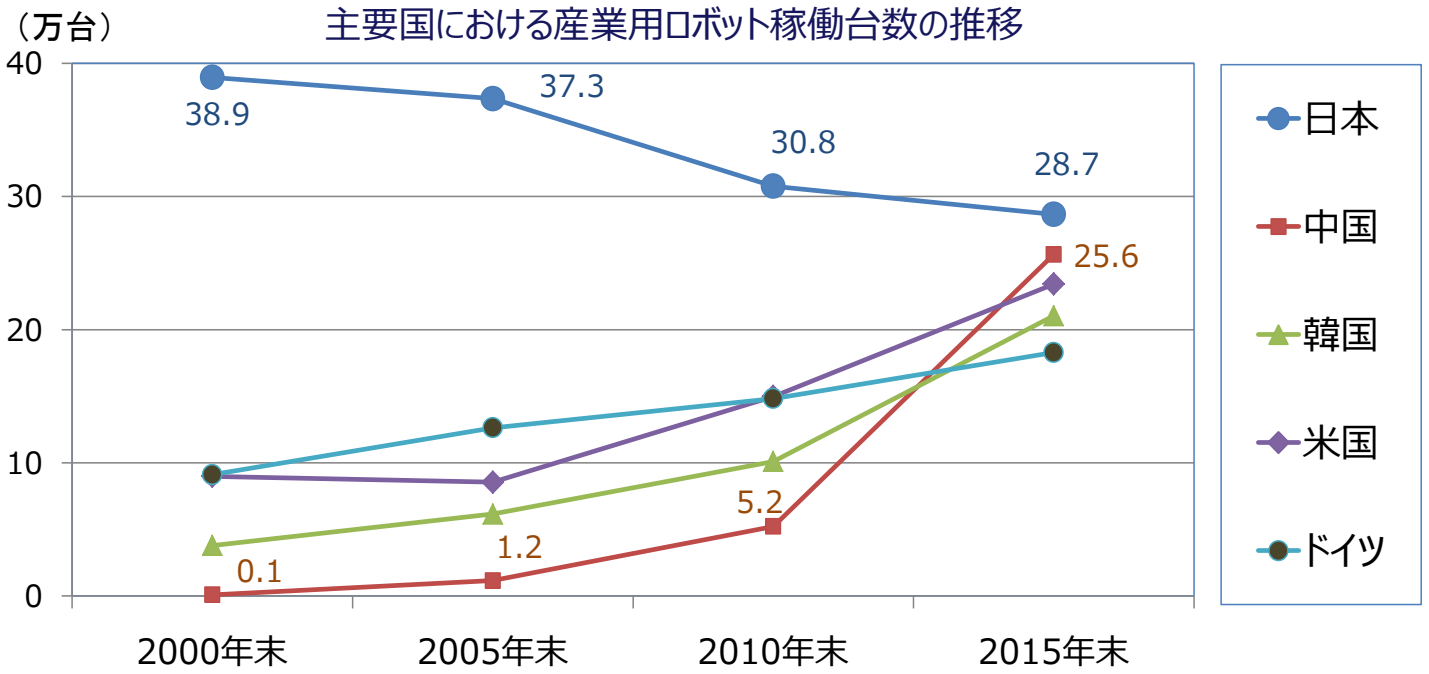


- ドローンは、「屋外用ロボット（フィールドロボット）」の一種ですが、特に広範な利活用が期待されているため、この講座では区別して紹介します。

産業用ロボットの普及と輸出

◆2015年末における日本の産業用ロボットの稼働台数は世界1位で、広く利用されています。

- 国際ロボット連盟（IFR）によれば、2015年末の産業用ロボットの稼働台数は1位が日本の28.7万台で、2位が中国25.6万台、3位がアメリカの23.4万台となっています。
  - 2015年末時点では日本における産業用ロボットの稼働台数は第1位ですが、第2位の中国は大幅に稼働台数を伸ばしています。
- 2015年における日本の産業用ロボットの出荷金額は、国内向けが2010億円、外国向けの輸出が4824億円と、輸出の方が多くなっています。
  - 日本で製造された産業用ロボットは過半が輸出されており、海外で稼働している日本製の産業用ロボットも数多くあることを示唆しています。



【出所】世界の産業用ロボット稼働台数［日本ロボット工業会］の公開資料に基づき作成
https://www.jara.jp/data/dl/stock-of-robot-2015.pdf

産業用ロボットの人工知能搭載とIoT化

◆産業用ロボットの人工知能搭載、IoT化によって、更なる効率化が期待できます。

- 工場全体の機器をインターネットに接続することで、外部からでも機器の稼働状況の可視化や最適化、制御を行うことができる**スマートファクトリー**と呼ばれる取り組みが進展しています。
- 先進的な産業用ロボットの制御として、ロボットが過去の成功や失敗の経験から学ぶ人工知能を搭載した学習ロボットが挙げられます。
  - ファナック株式会社では、産業用ロボットが学習することで、学習前よりも10%作業時間を短縮できる事例を示しています。
- ロボットがIoT化することによって、一つのロボットの学習を他のロボットに伝達・共有することができます。

学習ロボットの作業効率化

【出所】ファナック株式会社 https://www.youtube.com/watch?v=DjPp3mM1RKM

IoT化によるマスカスタマイゼーション

◆産業用ロボットのIoT化は、マスカスタマイゼーションを進展させます。

- 産業用ロボットのIoT化の進展によって、多品種少量生産の**マスカスタマイゼーション**が進展します。
  - マスカスタマイゼーションは、大量生産を意味するマスプロダクションと、顧客の個別要望に応じた生産、サービスを行うカスタマイゼーションの合成語です。
- ZOZOTOWNでは、センサーが付属するZOZOSUIT（ゾゾスーツ）を配付し、得られたデータを利用することで、個々人の身体のサイズに合った服を販売しています。
  - ZOZOSUITは、利用者の身体に合わせて伸縮し、15,000箇所の身体のサイズを測ることができるとされています。
- 島精機製作所では、コンピュータ制御の自動編機を開発・販売しています。
- ホールガーメントと呼ばれる縫製不要の衣服が、自動織機から立体的に編み上がります。

伸縮センサー内蔵したZOZOSUIT

【出所】ZOZOSUIT［株式会社スタートトゥデイ］
http://zozo.jp/zozosuit/

コンピューター制御で作られるホールガーメント

【出所】ホールガーメントのメリット［島精機製作所］
http://www.shimaseiki.co.jp/wholegarment/

産業用ロボットと人間の協働

◆ロボットと人間は、より得意な分野の作業を分担する形で協業することができます。

- 経済産業省が公表している「ロボット活用の基礎知識」では、ロボットの活用によって、**「貴重な人材を付加価値が高くクリエイティブな仕事に注力させることができる」**と紹介しています。

【出所】ロボットの活用に関する一般向けコンテンツを作成しました［経済産業省］
http://www.meti.go.jp/press/2016/03/20170329001/20170329001.html

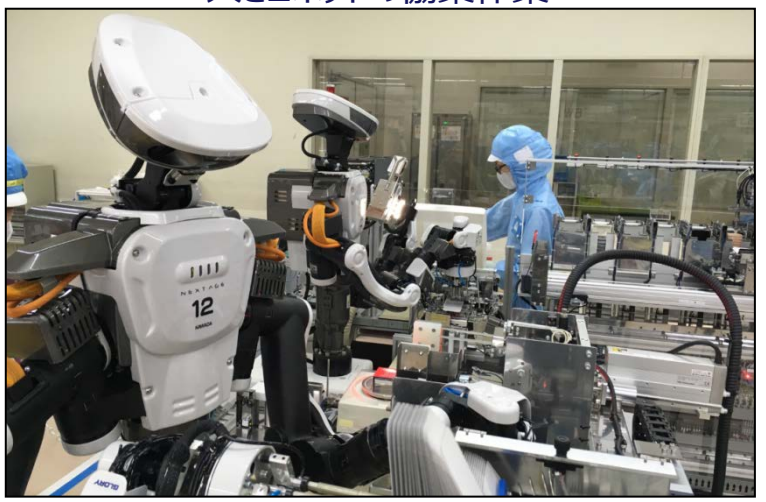
- カワダロボティクス株式会社では、人間と協業できる双腕型産業用ロボットやヒト型ロボットを製作、販売しています。
  - 双腕型産業用ロボットのNEXTAGE（ネクステージ）は、近くで作業を行う人間の安全に配慮した低出力モーターによる構成など、人間との共同作業に配慮して設計されています。

双腕型産業用ロボットのNEXTAGE

【出所】NEXTAGE［カワダロボティクス株式会社］
http://nextage.kawada.jp/

- 資生堂は2017年3月より静岡県掛川工場に産業用人型ロボットを導入し、作業者1名、ロボット2台でチームを組む形で製品を仕上げています。
  - ロボットはケースの組み立てやレーベルの貼り付けを行い、作業者は検品作業等、現在の産業用ロボットでは代替できない業務に専念します。

人とロボットの協業作業



【出所】組立工程に人型ロボットを試験導入［資生堂］
https://www.shiseidogroup.jp/news/detail.html?n=00000000002150

医療用ロボット

◆医療分野では手術の実施や支援を行うロボットの活用が進んでいます。

- アメリカのIntuitive Surgical社が発売している手術支援ロボット「da Vinci（ダ・ヴィンチ）」は執刀医が患者から離れた位置で、3D画像を確認しながらロボットアームで手術を行います。
  - ダ・ヴィンチは、2016年9月末において、237台が国内の医療機関に導入されています。
  - ダ・ヴィンチのロボットアームに取り付ける鉗子は40種類以上があり、様々な外科手術を実行可能です。

手術支援ロボットda Vinci（ダ・ヴィンチ）

【出所】da Vinci［日本ロボット外科学会］
http://j-robo.or.jp/da-vinci/index.html

- 東京工業大学と東京医科歯科大学の両大学発のベンチャー企業リバーフィールド株式会社が開発したEMARO（エマロ）は、執刀医の頭部の動きと連動して動作する内視鏡ホルダーロボットです。
  - 内視鏡（腹腔鏡）手術は、患者の腹部に複数の小さな穴を開け、その穴から小型カメラの内視鏡や手術器具を入れて手術を行います。
  - 一般にスコピストと呼ばれる内視鏡の操作担当者が必要ですが、EMAROによって執刀医が一人でも内視鏡手術が可能となります。

内視鏡ホルダーロボット EMARO（エマロ）

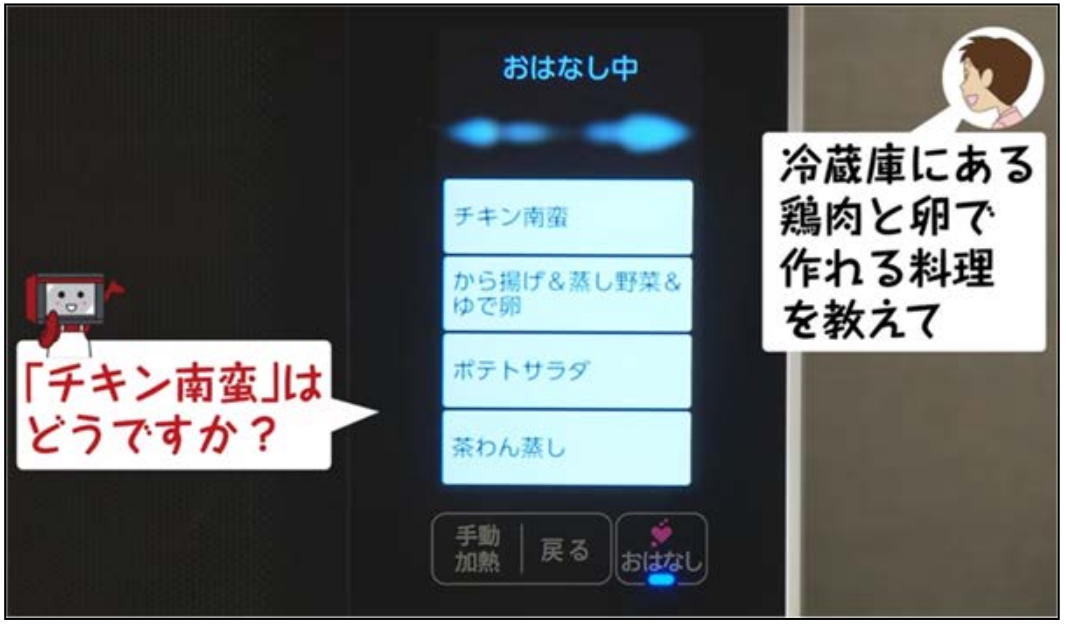
【出所】EMARO［リバーフィールド株式会社］
https://www.riverfieldinc.com/product/

人工知能を内蔵した家電製品（ロボット家電）

◆ロボット家電とも呼ばれる人工知能が組み込まれた家電製品も発売されています。

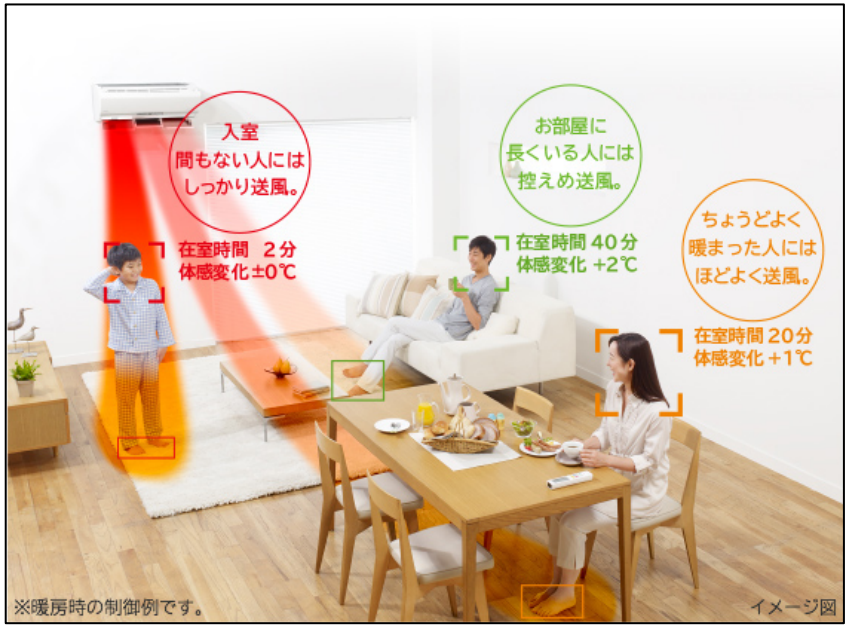
- シャープ株式会社が販売しているオーブンレンジ「ヘルシオ」では音声を認識し、人工知能が材料、季節、調理履歴などを考慮して献立を提案します。
- 株式会社日立製作所が販売しているエアコンでは、カメラと人工知能によって身体の部位、温度、湿度を把握して、屋内の場所を狙った温度管理ができます。

音声認識した材料での料理を提案するオーブンレンジ



【出所】オーブンレンジ「ヘルシオ」［シャープ株式会社］
https://www.youtube.com/watch?v=K6DmU7eXjzA

体感温度に合わせた温度調整をするエアコン



【出所】くらしカメラAIを搭載したエアコン［株式会社日立製作所］
https://kadenfan.hitachi.co.jp/ra/lineup/xseries_h/feature04.html

□ これらの家電製品はインターネットと接続していることでIoTであるとともに、「センサー」「人工知能」「（レンジ、エアコンなどの）駆動系」があることで、ロボットと考えることもできます。

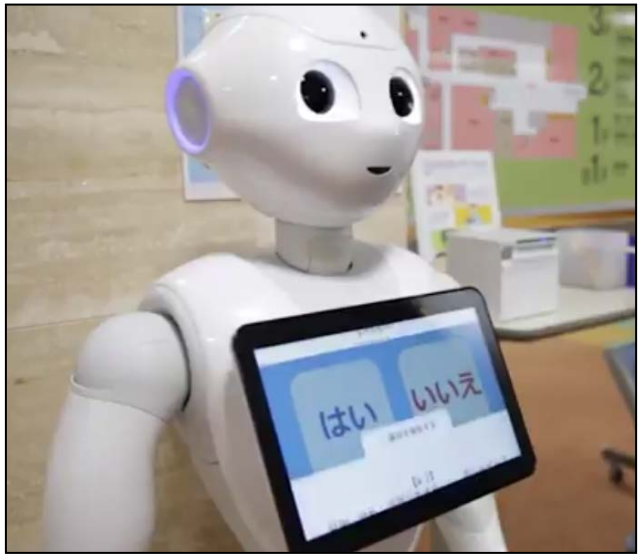
コミュニケーションロボット

◆会話などによって、人間とコミュニケーションがとれるロボットは広範に活用されています。

- 会話などによって、人間とコミュニケーションがとれるロボットを**コミュニケーションロボット**といいます。

- ソフトバンクロボティクス株式会社が開発しているPepper（ペッパー）は人間の感情を判断し、自らの感情を表現するコミュニケーションロボットです。
- Pepperに質疑応答を設定することで、様々な産業で活用されています。
  - Pepperは月額制で借りることができ、2017年3月末時点で2000社超の企業で利用されています。

病院で利用されるPepper



【出所】Pepper［ソフトバンクロボティクス株式会社］
https://www.softbank.jp/robot/special/biz/

- 株式会社タカラトミーから発売されているCOZMO（コズモ）は、付属のキューブを使って人間とゲームができ、自らの感情を表現するロボットです。
- ブロックを連ねる形で、COZMOの動きをプログラミングすることもでき、プログラミング教育にも有益です。

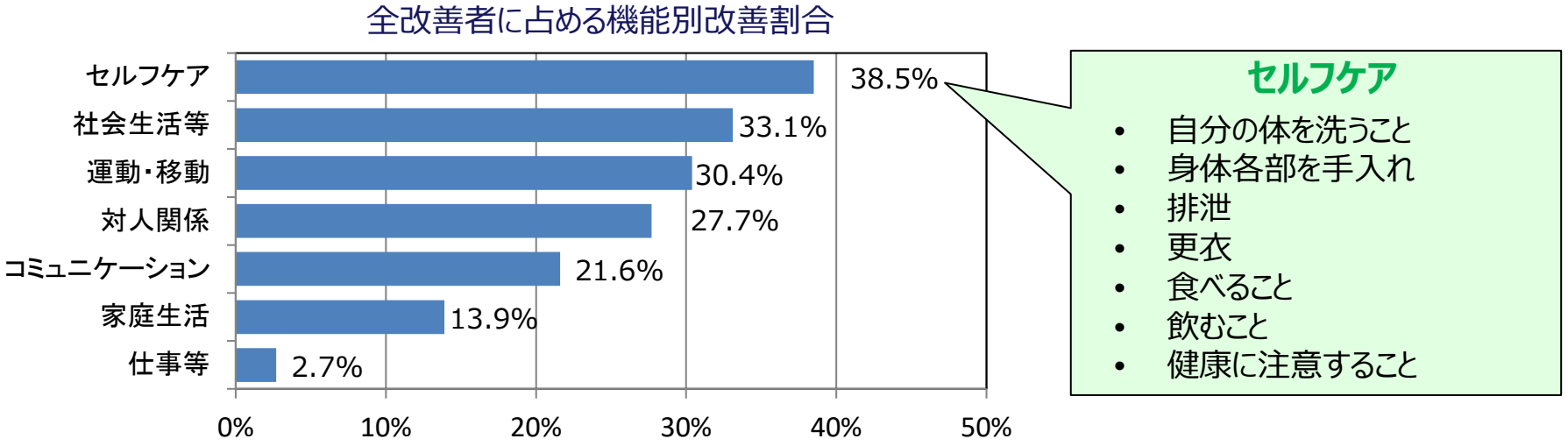
プログラミングが可能で人間とゲームができるCOZMO

【出所】COZMO［株式会社タカラトミー］
http://www.takaratomy.co.jp/products/cozmo/

介護分野におけるコミュニケーションロボットの活用

◆コミュニケーションロボットは、介護分野における活用も期待されています。

- 国立研究開発法人 日本医療研究開発機構は、コミュニケーションロボットの大規模実証実験として、介護分野における活用を調査しました。
- コミュニケーションロボットによる被介護者の心身の機能改善を調査すると、有効被験者866人のうち34.2%に相当する296人に何らかの改善が見られました。
  - 心身の機能は、ICF（国際生活機能分類）と呼ばれる基準にて分類され、調査されました。
- 機能別の改善割合が最も高かったのは「セルフケア」であり、「コミュニケーション」の改善割合を超えていました。
- 調査結果は、コミュニケーションロボットが、被介護者の「コミュニケーション」の機能のみならず、「セルフケア」をはじめとする他の心身の機能を改善させることを示しています。



【出所】介護分野におけるコミュニケーションロボットの活用に関する大規模実証試験報告書［日本医療研究開発機構|大川弥生］
http://robotcare.jp/wp-content/uploads/2017/07/communi_robo_veri_test_report.pdf

農業用ロボット・建機ロボット

◆農業分野、建築分野でもロボットの活用が始まっています。

- パナソニック株式会社は、トマトの赤色を自動的に判別し、熟れた実だけを自動で収穫するトマト収穫ロボットの発売を目指しています。

熟れた実を判別して採取するトマト収穫ロボット

【出所】トマト収穫ロボット［パナソニック株式会社］
https://news.panasonic.com/jp/stories/2017/52916.html

- コマツのICT油圧ショベルは、位置情報と施工設計の3次元データを照合して、自動で設計図に合わせた土木工事が可能です。

施工設計データに基づき稼働するICT油圧シャベル

【出所】ICT油圧シャベル［コマツ|株式会社小松製作所］
http://smartconstruction.komatsu/introduction/ictkenki.html

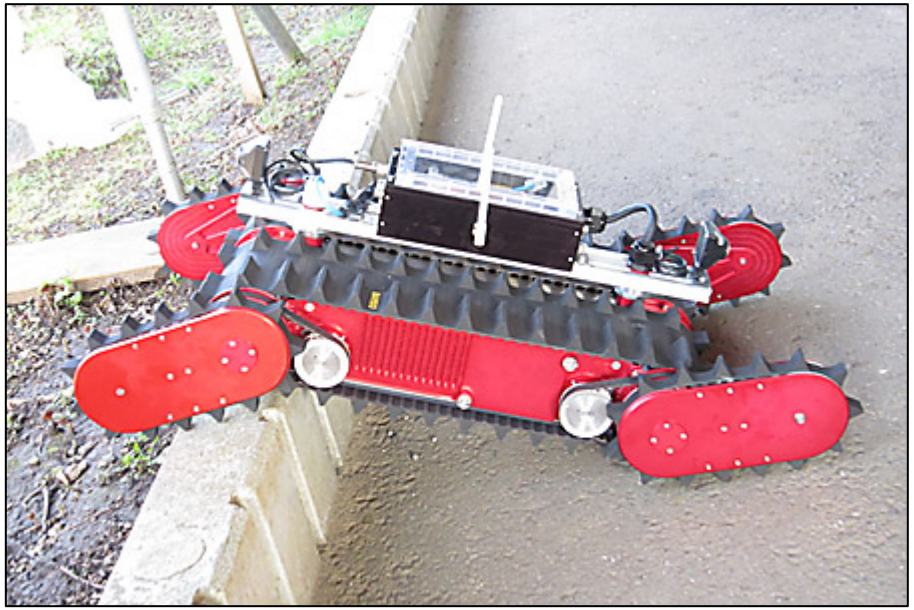
□ 農業分野、建設分野では、今後の熟練労働者の不足を補うロボットの開発が進められています。

災害対応ロボット、運搬ロボット

◆遠隔操作ができる災害対応ロボット、人に追従して運搬を行う追従運搬ロボットがあります。

- 株式会社移動ロボット研究所が発売している災害対応ロボットのアルバトロスは遠隔操作が可能で、温度やガス濃度を検知し、災害現場等における情報収集を行うことができます。

災害現場対応ロボット（アルバトロス）



【出所】災害対応ロボット［株式会社 移動ロボット研究所］
http://www.irobo.co.jp/

- 株式会社 タナックスが販売しているサウザーは、自動追従、衝突回避、無人ライン走行の3つの機能を持つ運搬ロボットです。
  - 無人ライン走行は、床面に貼られた専用テープに沿うルートを走行します。

追従運搬ロボット（サウザー）

【出所】サウザー［株式会社タナックス］
http://tana-x-thouzer.com/

□ 遠隔操作で実力を発揮するロボットもあれば、人に追従することで実力を発揮するロボットもあります。

移動用ロボット

◆自動車においても、人工知能が導入され移動用ロボットとしての活用が始まっています。

- 移動用ロボットとして、インターネットと通信を行う車の**コネクティッドカー**や自動車への人工知能搭載が挙げられ、**官民ITS構想**では、運転支援から完全自動運転への段階的な進展が見込まれています。
  - ITSは、「Intelligent Transport Systems（高度道路交通システム）」を略しており、官民ITS構想の各段階は、講座4-4(8)移動で紹介します。

【出所】官民 ITS 構想・ロードマップ 2017［高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議］
https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20170530/roadmap.pdf

- 日産自動車株式会社が発売している電気自動車LEAF(リーフ)では、カメラとソナーから得られた情報を人工知能が利用することで、運転席にあるボタンを押すことで適切な場所に駐車することができます。

電気自動車リーフにおけるカメラ・ソナーの位置

【出所】プロパイロット パーキング［日産自動車株式会社］
https://www.nissan-global.com/JP/TECHNOLOGY/OVERVIEW/propilot_park.html

- トヨタ自動車株式会社が「Concept-愛i（コンセプト・アイ）」シリーズを開発しており、人工知能を搭載した自動車が人の感情を理解し、パートナーになることを目指しています。

Concept-愛i（コンセプト・アイ）シリーズ

【出所】Concept-愛i［トヨタ自動車株式会社］
http://www.toyota.co.jp/jpn/tech/smart_mobility_society/concept-i/

□ 講座1-2で紹介したロボットスーツも、人間の移動を支援するケースがあり、移動ロボットとして考えるケースがあります。

ドローンの定義

◆様々な用途での活用が期待される屋外用ロボットとして、ドローンが挙げられます。

- **ドローン（Drone）**は、遠隔操作や自動操縦によって飛行する無人航空機の総称です。
  - 自動操縦機能を持つことをドローンの定義に含めるケースと含めないケースがあり、含めないケースではラジコンのヘリコプターもドローンに含まれます。
- 一般にドローンは、3つ以上の回転翼を搭載しているマルチコプター型の無人航空機を指しています。
  - 一般的なドローンは、回転翼の数が多いことに加えて、加速度センサーやジャイロセンサーによる情報から自動的に姿勢制御をするため、安定した飛行が可能になっています。
- センサー等から得られた情報に基づき、人工知能等による制御が可能なドローンは、屋外用ロボットの一種と考えることができます。
- ドローンは立ち入り困難な地域でのカメラ撮影や遠隔地への運搬、農業での利用、災害救助、測量など様々な用途が期待されています。

撮影用ドローン

運搬用ドローン

農薬散布用ドローン

ドローンの急速な普及

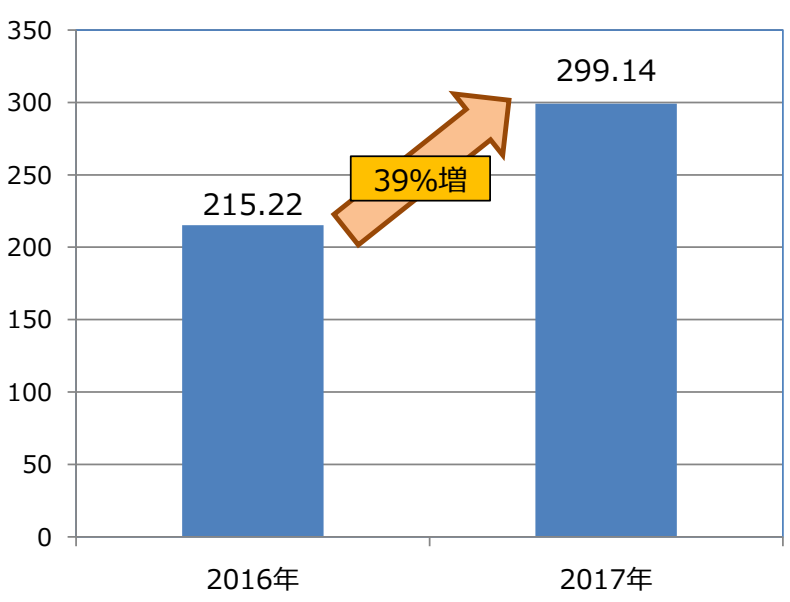
◆ドローンは、世界においても、日本国内においても、急速な普及が見込まれています。

- ドローンの活用は、世界においても、日本においても注目されており、出荷量や市場規模が推計されています。

- 米国の市場調査会社Gartner（ガートナー）が、2017年2月に発表したレポートによれば、2017年における民間用ドローンの世界出荷台数は2016年比39%増の299万台と推計されています。

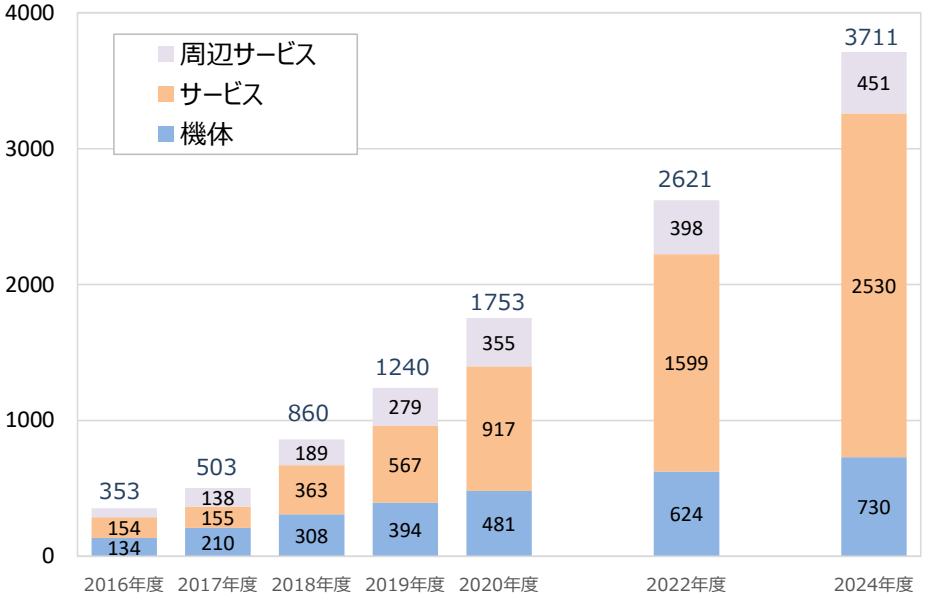
- インプレス総合研究所が2018年3月に発表した国内ドローンの市場規模予測によれば、2016年度の353億円から2024年度の3711億円まで8年間で10倍以上の市場規模拡大が見込まれています。

世界のドローン出荷台数の推計値



【出所】Gartner
https://www.gartner.com/newsroom/id/3602317

国内のドローンビジネスの市場規模の予測



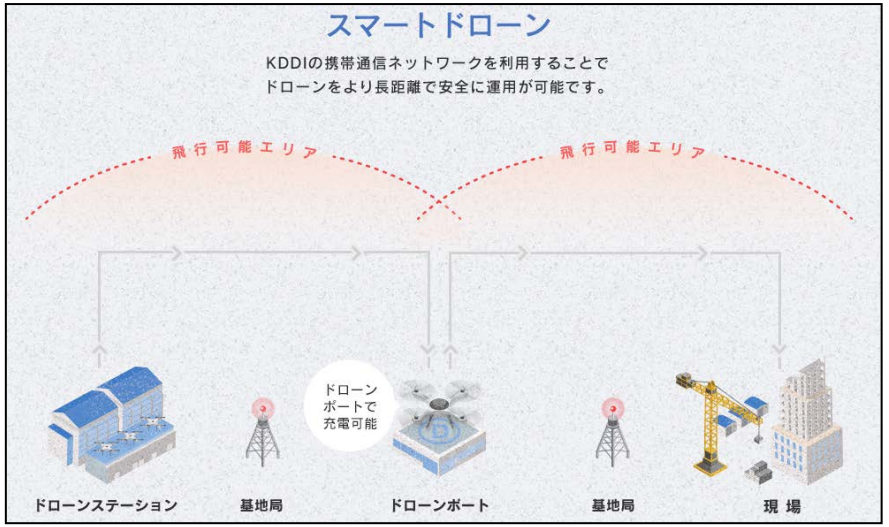
【出所】ドローンビジネス調査報告書2018
［インプレス総合研究所］のニュースリリースに基づき作成
https://www.impress.co.jp/newsrelease/2018/03/20180315-01.html

ドローン活用のプラットフォーム

◆モバイル通信事業社を中心にドローン活用のプラットフォームが形成されています。

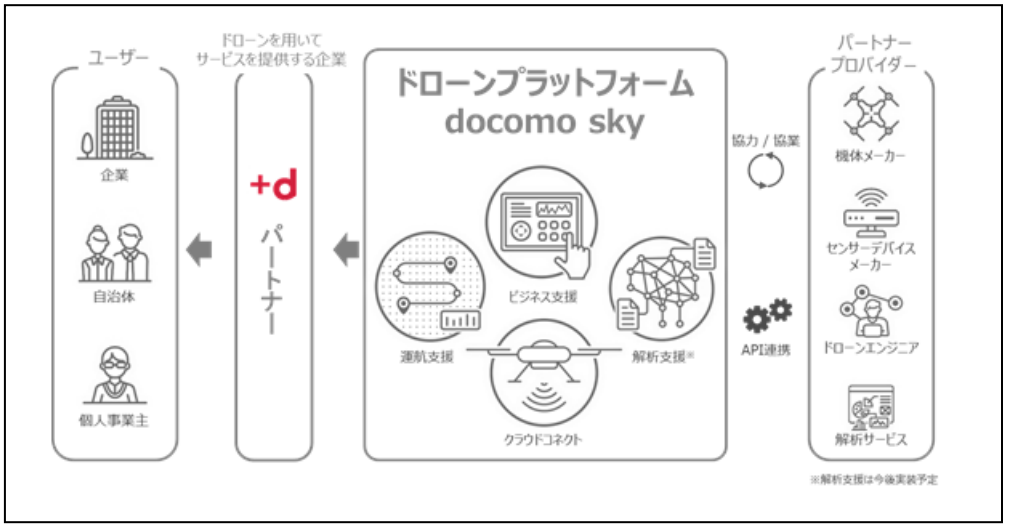
- ドローンの利用者は、KDDIのスマートドローンプラットフォームを利用することで、携帯通信ネットワークを利用することで、遠隔制御による安全な長距離飛行が可能となります。
- 2017年11月にKDDIは、「LTEのモバイル通信ネットワーク」「3次元地図」「ドローンポート」を用いた約6.3kmの完全自律飛行に成功しました。
- 2018年2月にNTTドコモは、ドローン運用をトータルサポートする「ドローンプラットフォーム docomo sky」を発表しました。
- 「ドローンプラットフォーム docomo sky」は、ドローンを用いたサービスを提供するパートナー企業向けにドローン運用を支援する仕組みです。

KDDIのドローンポートにおける充電イメージ



【出所】スマートドローン［KDDI株式会社］
http://smartdrone.kddi.com/

「ドローンプラットフォーム docomo sky」の概念図



【出所】ドローンプラットフォーム docomo sky［株式会社NTTドコモ］
https://www.nttdocomo.co.jp/info/news_release/2018/02/21_00.html

ドローンに関する規制と利活用

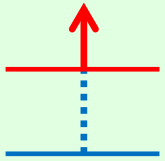
◆ドローンの利活用と安全確保の間でのバランスを検討する必要があります。

- 2015年に航空法の一部が改正され、ドローン等の無人航空機の飛行ルールが定められました。
  - 改正された航空法以外にも、小型無人機等飛行禁止法、電波法などにも、ドローンの飛行に関連する規制があります。
- 国土交通省のウェブサイトでは、ドローンの飛行ルールを示すとともに飛行許可を得るための手続きを示しています。
  - ドローンの飛行可否を示す地図がウェブ上に公開されており、右下図に首都圏における飛行禁止区域を示しています。地図上に濃い色が塗られているエリアは、屋外における150m以下のドローンの飛行に許可が必要です。

【出所】無人航空機（ドローン・ラジコン機等）の飛行ルール［国土交通省］
http://www.mlit.go.jp/koku/koku_tk10_000003.html

許可なく200g以上の
ドローンを屋外で飛ばせない場所

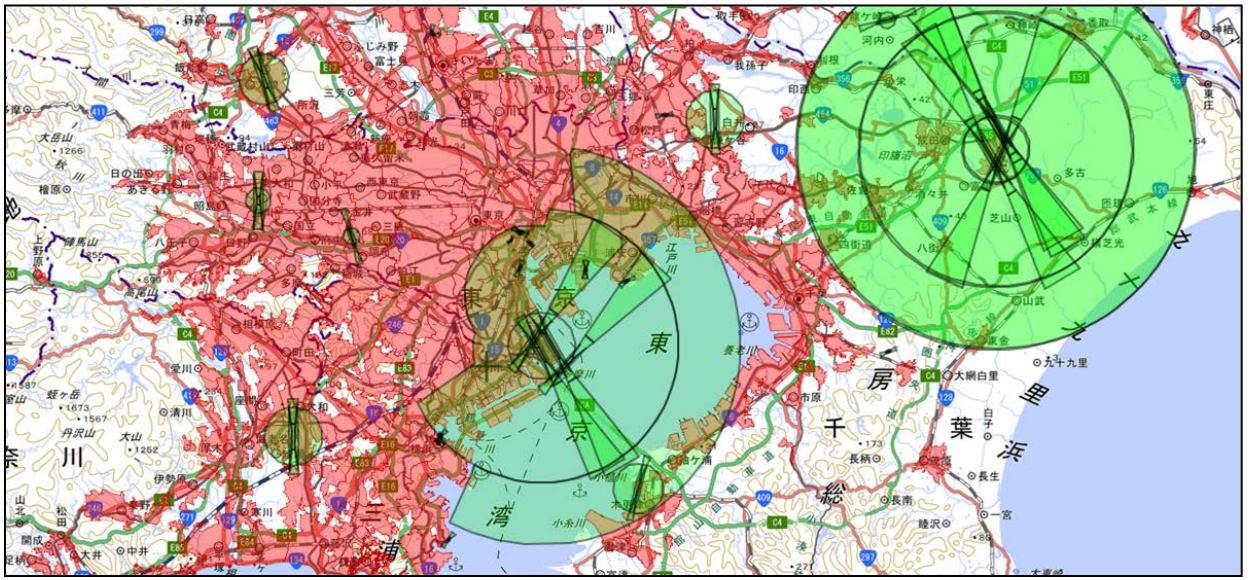
空港周辺（進入表面等）



高度150m以上

人口集中地区

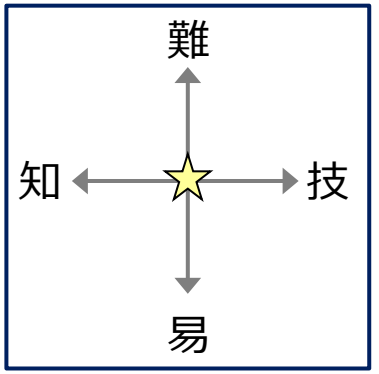
ドローンの飛行可否を示す地図（首都圏）



【出所】無人航空機の飛行の許可が必要となる空域について［国土交通省］からリンクされた国土地理院地図 http://www.mlit.go.jp/koku/koku_fr10_000041.html

◘ ドローンの活用は法整備を踏まえて、利便性の安全確保の間のバランスをとることが求められています。

総務省　ICTスキル総合習得教材



［コース1］データ収集

1-5：APIによるデータ収集と利活用

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_1_5.pdf

	1	2	3	4	5
［コース1］データ収集					●
［コース2］データ蓄積					
［コース3］データ分析					
［コース4］データ利活用					

本講座の学習内容［1-5：APIによるデータ収集と利活用］

【講座概要】

- APIの基本的なイメージを紹介し、APIの分類を示します。
- 近年注目を集めているAPIとして、オープンAPIを事例を挙げて紹介します。
- APIを活用したマッシュアップやAPIエコノミー、APIとIoTの親和性を説明します。
- オープンAPI以外とは異なる用途のAPIとして、デバイスWeb APIを紹介します。
- APIで一般に利用されるファイル形式（XML、JSON）の特性を示します。

【講座構成】

座学

- [1] APIの用途と分類
- [2] オープンAPIの事例
- [3] オープンAPIの特性とIoTとの親和性
- [4] デバイスWeb API
- [5] Web APIとファイル形式

【学習のゴール】

- ✓ APIの基本イメージと分類を理解する。
- ✓ 様々なオープンAPIの活用事例を紹介できる。
- ✓ APIに関連するマッシュアップ、APIエコノミーの意味を理解する。
- ✓ APIで一般に利用されているファイル形式（XML、JSON）の特性を把握する。

APIによる情報収集

◆この講座では、「他者が提供している情報を収集する」という観点から、APIを説明します。

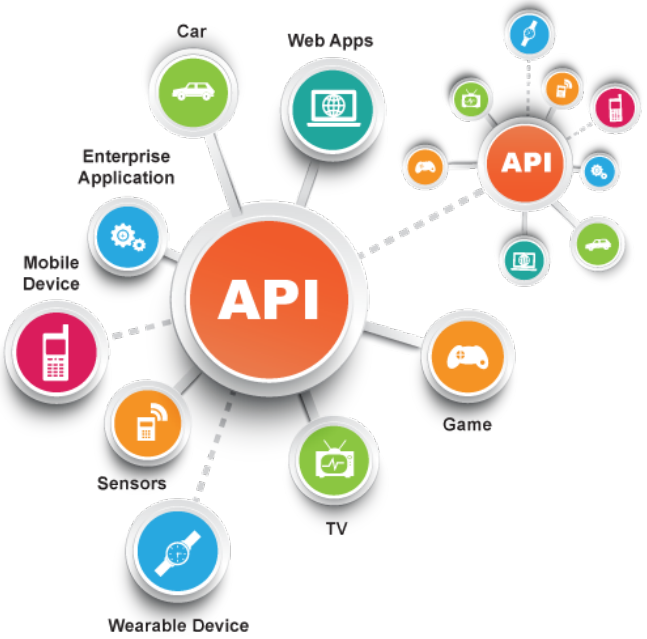
- 他者が提供する情報を収集したり、提供している機能を利用する方法として、**API（エーピーアイ：Application Programming Interface：アプリケーション プログラミング インターフェイス）**が挙げられます。

- APIの最後のIのInterfaceは、「接点」「接続面」と訳され、「つなぐもの・媒介物」を意味し、APIはプログラムによって「様々な機器やアプリケーション（ソフトウェア）をつなぐもの」というイメージです。
  - Interfaceはハードウェア面では「機器の接続部分」、ソフトウェア面では情報交換の「（受付部分の）ルール」を指すことが一般的です。

- APIには様々な種類がありますが、近年においては**Web API**が特に注目を集めています。
  - Web APIは、外部機器との通信を行い、APIを通じた情報交換がHTTPと呼ばれるウェブサイトの閲覧と同じ通信方式を利用しているため、前にWebとついています。
  - Web API以外のAPIとして、Windowsの機能をプログラムから利用するためのルールとしての「Win64 API」が挙げられます。こうしたAPIの利用においては、インターネット通信やWebの利用は不要です。

- 近年、APIの前に特定の語句を付けない場合でも、Web APIを指すケースが多く、この講座ではWeb APIに焦点を当てて説明をします。
  - 以降ではこの講座においても、APIの前に「Web」と付けないものはWeb APIを指しています。

APIの「つなぐイメージ」



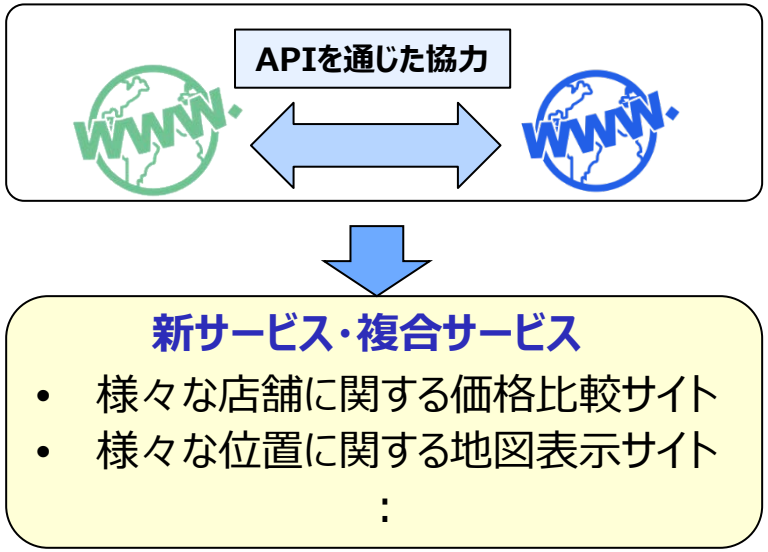
□ Web APIは、インターネット接続を前提としており、IoTとの親和性が高くなっています。

コンピュータ（プログラム）同士で協力を行うAPI

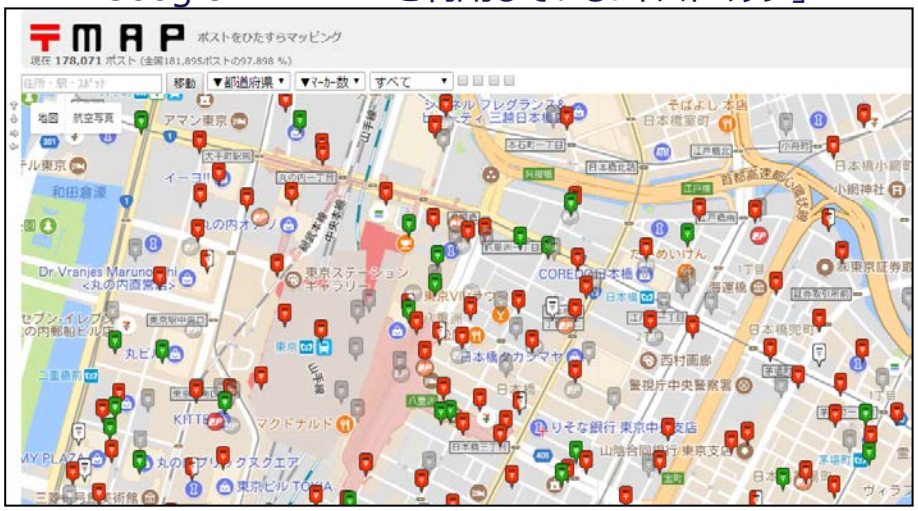
◆APIは「コンピュータ（プログラム）同士が協力する機能・仕組み」を指しています。

- APIを利用すると、複数のコンピュータ（プログラム）で情報交換・協力することで、単体のコンピュータ（プログラム）では、できなかったことができるようになります。
  - 協力すれば大きなこと、新しいことができるようになるのは人間も、コンピュータ（プログラム）も同様です。
- Web APIを利用することで、他者が提供している情報やインターネット上のサービスを自分のウェブサイトや自作アプリに利用することもできます。
  - 人間がプログラムを作成することで、API経由でコンピュータ（プログラム）に情報を与えたり、情報を引き出したりすることもできます。
- 「ポストマップ」は、Google MapのAPIを利用することで、Google Map上に郵便ポストの位置を表示しています。
  - ポストマップでは、APIを通じてGoogle Mapのサービスを借りてくることで、地図表示は専門サービスのGoogle Mapに任せて「ポストマップ」の付加価値である郵便ポストの位置指定に注力することができます。

APIのウェブ利用による新たなサービス



Google MAP APIを利用している「ポストマップ」

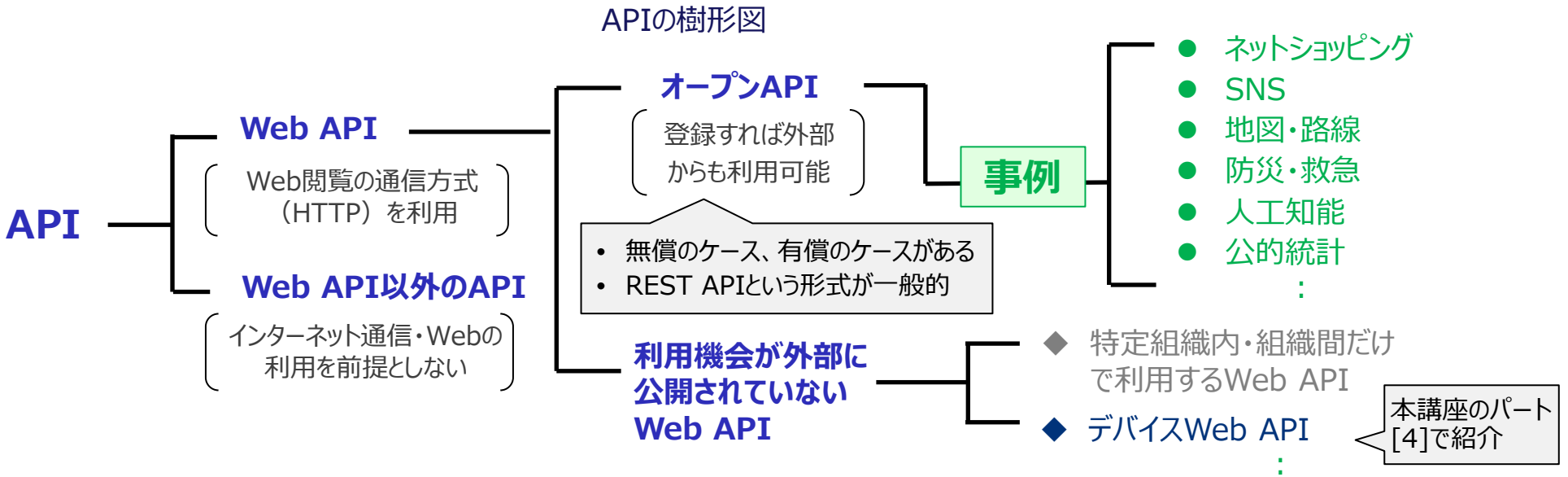


【出所】ポストマップ［30maps.com］
https://www.postmap.org/

オープンAPI

◆オープンAPIと呼ばれる無関係な外部組織を含めて利用することができるAPIがあります。

- 無関係な外部組織を含めて、第三者による利用が可能なAPIは**「オープンAPI」**と呼ばれます。
  - オープンAPIは技術的にはREST APIまたはRESTful APIと呼ばれる「パラメータを指定して特定のURLにHTTPでアクセスする形式、過去の情報に依存せず各アクセスで要求が完結する形式」をとっているケースが一般的です。
  - オープンAPIとは対照的に一般公開されず、同一組織内やあらかじめ定められた組織間でのみ利用可能な内部用のAPIもあります。
- 利用に応じた課金があるオープンAPIもあれば、無償で利用することができるオープンAPIもあります。
  - 無償利用ができるAPIであっても、一般にAPIを利用するためには事前登録が必要で、APIの提供者はAPIの利用者ごとの利用状況が把握できます。
  - Google MAPのウェブ表示用のAPIは、1日25,000回までのマップロードは無料ですが、それ以上はマップロード1,000回あたり0.5$の課金があります。



□ 続いてオープンAPIとして、「ネットショッピング」「SNS」「地図・路線」「防災・救急」「人工知能」「公的統計」に関する事例を示します。

ネットショッピングに関するAPI事例

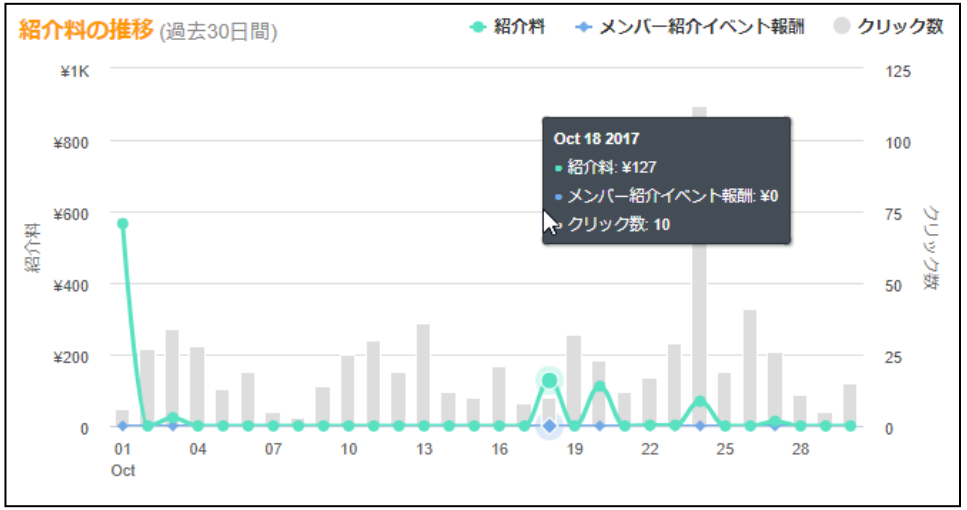
◆Amazonや楽天といったネットショッピングサイトではAPIを公開しています。

- ネットショッピング関連のAPIでは、アフィリエイト機能によって、APIの利用者が収入を得られるケースもあります。
  - アフィリエイトとは、ネットショッピングやウェブ広告においては、広告主の商品やサービスを紹介し広告主に収入があった場合に、紹介者が紹介料を受け取ることができる仕組みを指しています。
  - アフィリエイト（affiliate）は、元来「提携、加盟」という意味があり、代わりに「仲間、共同事業」を表すアソシエイト（associate）と呼ぶこともあります。

- AmazonのProduct Advertising APIにおいては、1日単位での販売額・紹介料の推移を把握できます。
  - Amazonは、クラウドのプラットフォームとしてAWS（Amazon Web Services）を運営していますが、Product Advertising APIと一般的なAWSは、現在は別のサービスだと考えてください。

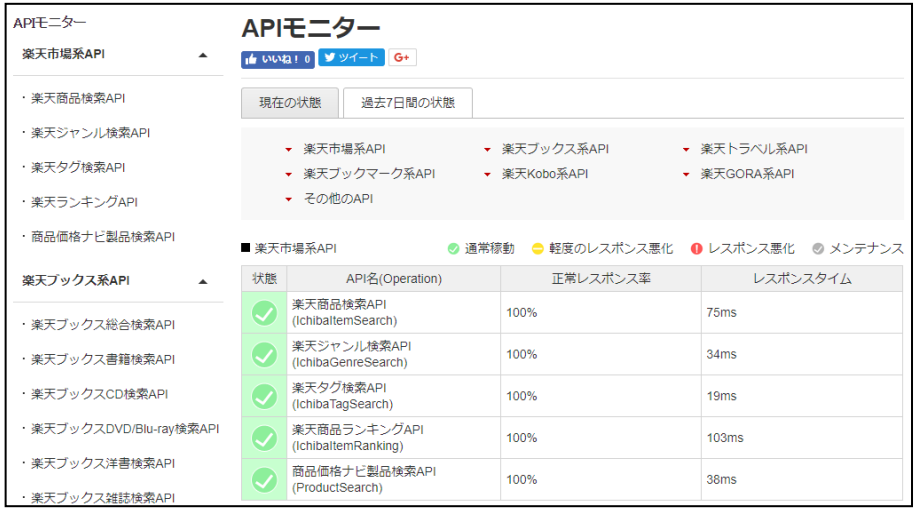
- 楽天のAPIモニターでは、各APIの稼働状況や反応時間を確認できます。
  - 楽天のAPIモニターでは、各APIの稼働の有無のみならず、正常レスポンス率の割合やマイクロミリ秒でのレスポンス時間が定量的に把握できるようになっています。

AmazonのProduct Advertising APIにおける紹介料表示



【出所】AmazonのProduct Advertising API［Amazon.com］
https://docs.aws.amazon.com/AWSECommerceService/latest/GSG/Welcome.html

楽天のAPIモニター



【出所】APIモニター［楽天株式会社］
https://webservice.rakuten.co.jp/status

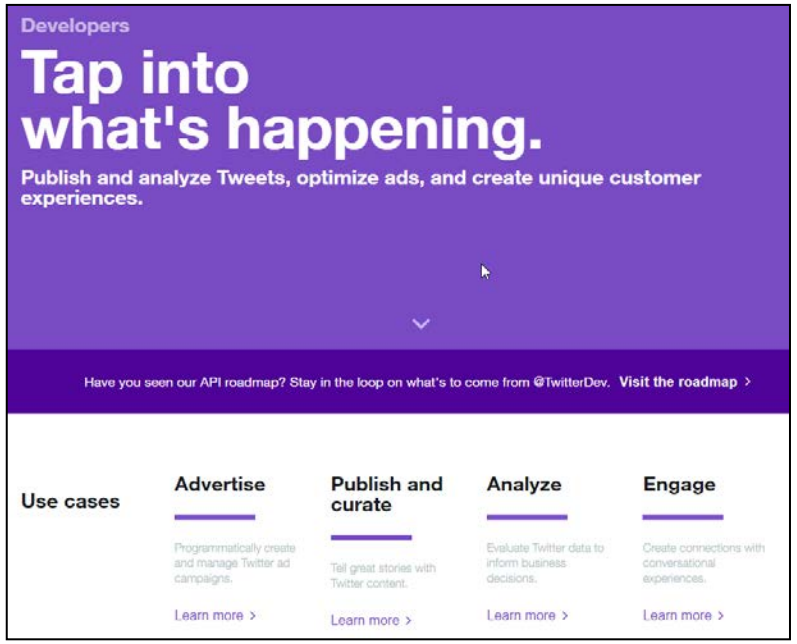
SNSに関するAPI事例

◆SNSのAPIを利用すると、公式サービス以外からの入力、出力のとりまとめができます。

- SNS（エスエヌエス）のAPIを利用すると、公式サービス以外からの情報投稿や情報の取りまとめが可能になります。
  - SNSは、「Social Networking Service」の略で、ウェブ上で共通の関心を持つ共同体形成や意見交換が可能なサービスを指しています。

- Twitter（ツイッター）のAPIを使うと、自サイトや自作アプリからTwitterにおけるツイート（Twitter上の発言）の検索や投稿が可能となります。

Twitter API



© 2018 Twitter, Inc.
https://developer.twitter.com/en.html

- TwitterのAPIを利用したTwilog（ついろぐ）では、日ごとにツイートをまとめたり、他のSNSとの情報を一括して表示することができます。

Twitter APIを活用したTwilog（ついろぐ）

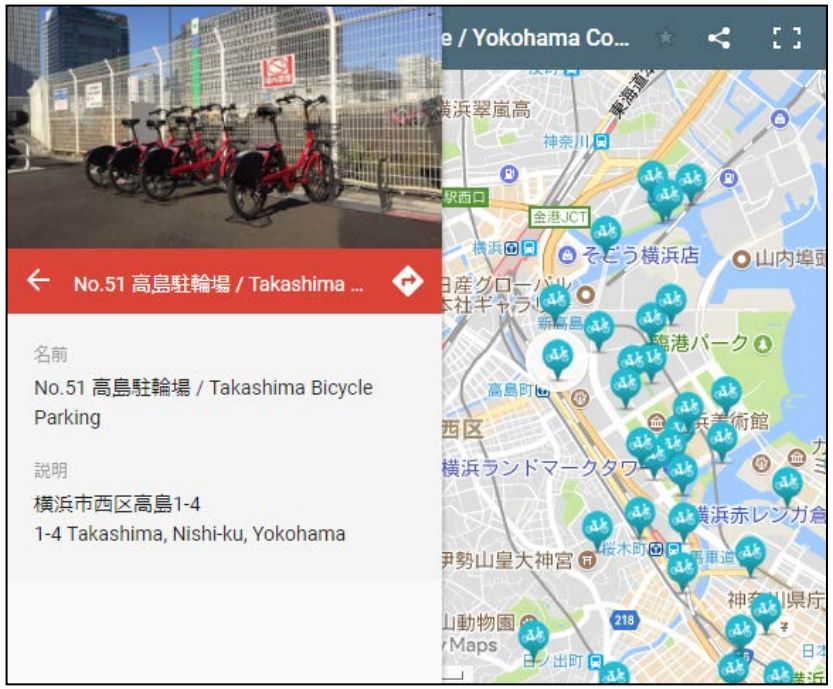
【出所】Twilog［ropross.net］
https://twilog.org/page-about

地図・路線に関するAPI事例

◆地図に関するAPIは、地図と位置以外にも追加情報を地図上に表すことができます。

- Google MapのAPIは、複数の地点を示すことも可能であり、横浜のコミュニティサイクルbay bikeのポートマップなど様々な位置表示に使われています。
  - bay bikeのポートマップでは、ポート（駐輪場）のアイコンをクリックすることで駐輪場の写真が表示されます。

横浜コミュニティサイクル　bay bikeの写真表示



【出所】bay bike［株式会社ドコモ・バイクシェア］
http://docomo-cycle.jp/yokohama/map/

- NAVITIME APIでは、移動手段別に指定地点からの移動可能な範囲を地図上に表示することができます。
  - 移動手段に応じた移動経路となっているため、移動手段によって移動可能な範囲の形状が異なります。

NAVITIME APIによる移動手段別の時間表示

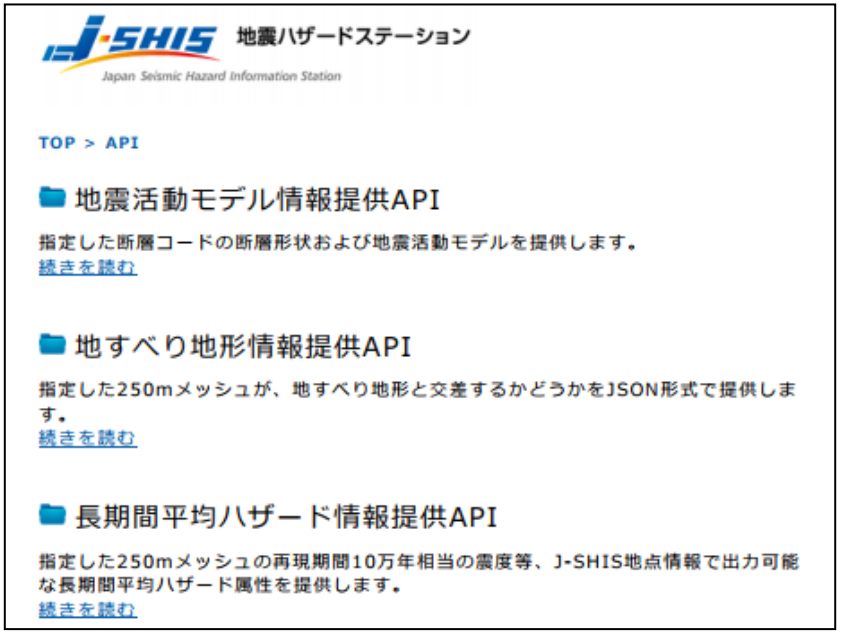
【出所】NAVITIME API［株式会社ナビタイムジャパン］
http://api-sdk.navitime.co.jp/api/

防災・救急に関するAPI事例

◆防災や救急に関するAPIは、小地域に関する防災、救急情報を確認することができます。

- 国立研究開発法人 防災科学技術研究所が提供するAPIでは、250mメッシュでの地震に関する情報が確認できます。
  - APIにて情報提供することで、防災科学技術研究所でしか集められない情報収集に注力し、利便性の高いウェブサイトやアプリの開発は、情報提供先の第三者に任せることができます。

防災科学技術研究所の地震情報に関するAPI



【出所】地震ハザードステーション［防災科学技術研究所］
http://www.j-shis.bosai.go.jp/category/opencat/api

- AEDオープンデータプラットフォームでは、自治体毎に様々な形式で公開されているAEDの情報を集約し、APIで公開しています。
  - AED（Automated External Defibrillator：自動体外式除細動器）は、心臓発作を起こした人を救う医療機器です。
  - サービスのタイトルに含まれる「オープンデータ」に関しては、講座4-1にて説明します。

AEDオープンデータ プラットフォーム

【出所】AEDオープンデータプラットフォーム
http://hatsunejournal.jp/w8/AEDOpendata/

人工知能に関するAPI事例

◆人工知能に関するAPIでは、画像認識、音声認識を行ったり、文章の分類が可能です。

- APIに送信するリクエストには、画像や音声を添付することができ、人工知能技術を活用したAPIに利用されます。
  - 人工知能や人工知能に含まれる分析技術の機械学習に関しては、講座3-5において説明をします。
  - このスライドでは、日本企業が提供する人工知能のAPIを紹介していますが、AWS、Microsoft、Google、IBMといった外資系企業も、この分野に注力しており、クラウドサービスと連携させる形で人工知能に関するAPIを提供しています。

- 株式会社NTTドコモが提供するAPIでは、画像認識、音声認識などの人工知能を利用した出力が得られます。

NTTドコモが提供するAPI

【出所】株式会社 NTTドコモ
https://dev.smt.docomo.ne.jp/

- 株式会社リクルートテクノロジーズが提供するAPI「A3RT（アート）」では、内容による文章の分類や画像の言語化や類似画像の検索ができます。

リクルートテクノロジーズが提供するA3RT（アート）

【出所】A3RT［株式会社リクルートテクノロジーズ］
https://a3rt.recruit-tech.co.jp/

公的統計・公的データに関するAPI事例

◆公的統計・公的データに関するAPIでは、対象を絞ってデータを収集することができます。

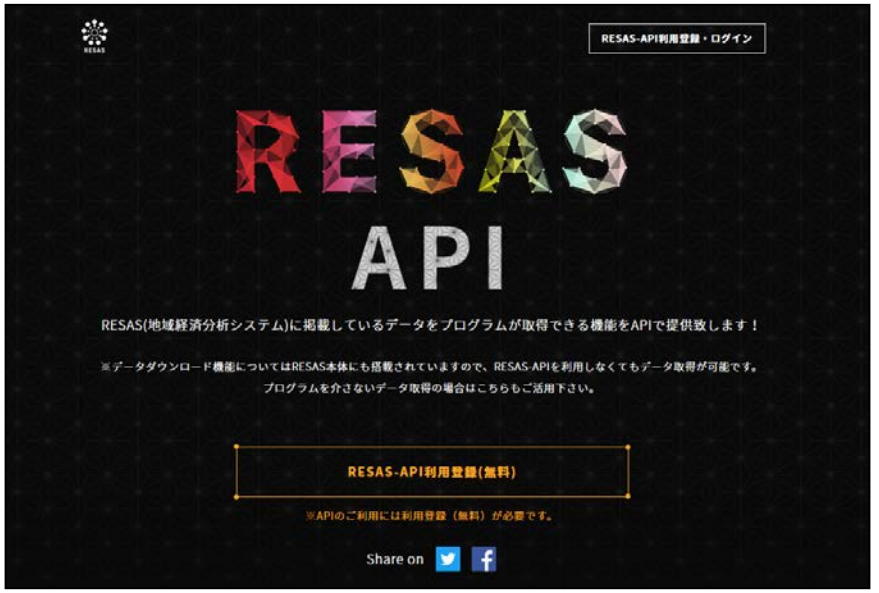
- 総務省統計局が運営するe-Stat APIでは、公的統計のデータを細目や期間を絞って収集できます。
  - 一般的なe-Statのウェブサイトでは、公的統計のデータを定型化された表形式でウェブブラウザに表示したり、Excelファイルをダウンロードしたりすることができます。
  - 講座4-3【参考2】では、Rのプログラミングを用いてe-Stat APIを利用する手順を具体的に示しています。

e-Stat API

【出所】e-Stat API［総務省 統計局］
https://www.e-stat.go.jp/api/

- 内閣官房（まち・ひと・しごと創生本部）が運営するRESAS APIでは、RESAS（リーサス）によって可視化されているデータの一部をAPIを通じて提供しています。
  - RESAS APIでは、将来人口の予測など、公的統計以外の公的データも提供しています。
  - 講座4-2では、RESASの操作方法や活用例を示しています。

RESAS API



【出所】RESAS API［内閣官房まち・ひと・しごと創生本部］
https://opendata.resas-portal.go.jp/

オープンAPIを利用したマッシュアップ

◆複数のAPIから得られた情報をマッシュアップで組み合わせ、新たなサービスが生まれます。

- 宮崎県の情報政策課が開発・公開している「ひなたGIS（地理情報システム）」は、RESAS APIをはじめとする様々なAPIの情報を集約した情報提供をしています。
  - 「ひなたGIS（地理情報システム）」は、2017年3月に内閣府が開催した第1回「RESASアプリコンテスト」で最優秀賞を受賞し、2017年5月に一般公開されました。

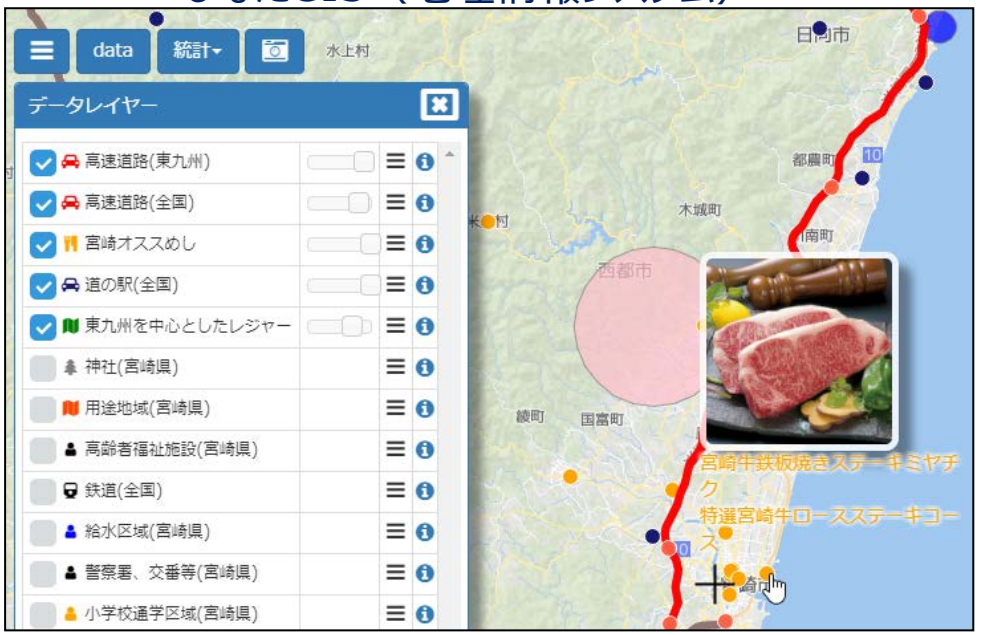
【出所】ひなたGIS（地理情報システム）の公開について（2017年5月）［宮崎県］
https://www.pref.miyazaki.lg.jp/johoseisaku/kense/joho/20170511004426.html

- 「ひなたGIS（地理情報システム）」は、
  - ◆RESAS-API
  - ◆e-Stat API
  - ◆気象庁から得た現在の気象情報
  - ◆高速道路・道の駅の情報
  - ◆宮崎オススめし、レジャー情報

  等の情報を組み合わせ、ウェブサイトに表示しています。

- 複数のウェブサービスや情報源の情報を組み合わせ、新たなサービスを作ることを**マッシュアップ**といいます。

ひなたGIS（地理情報システム）



【出所】ひなたGIS［宮崎県］
https://hgis.pref.miyazaki.lg.jp/hinata/

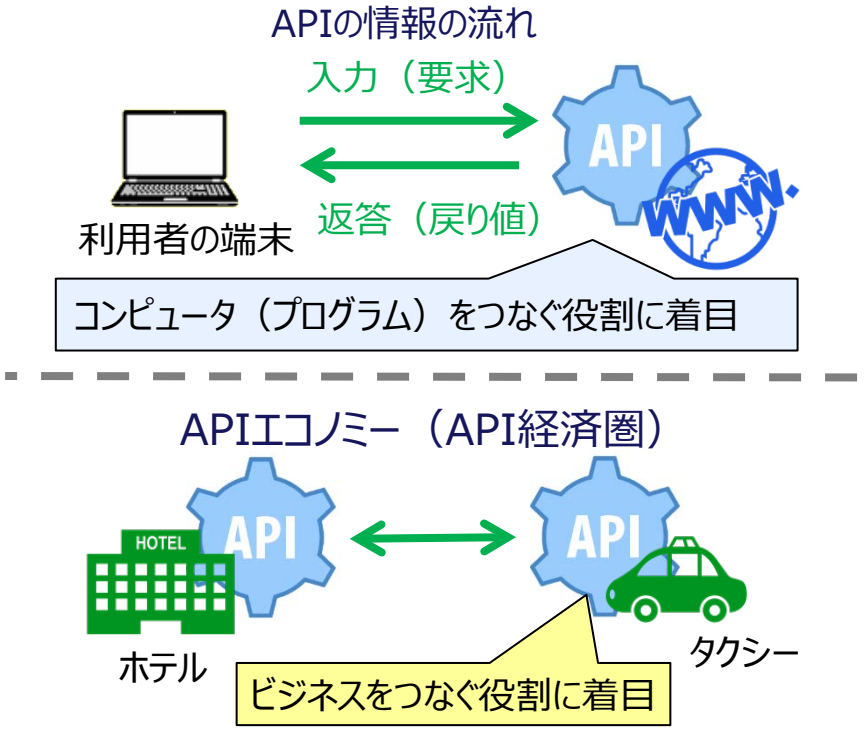
□ APIは既存サービスの利便性を高めるのみならず、情報・機能を組み合わせ新たなサービスを創造することができます。

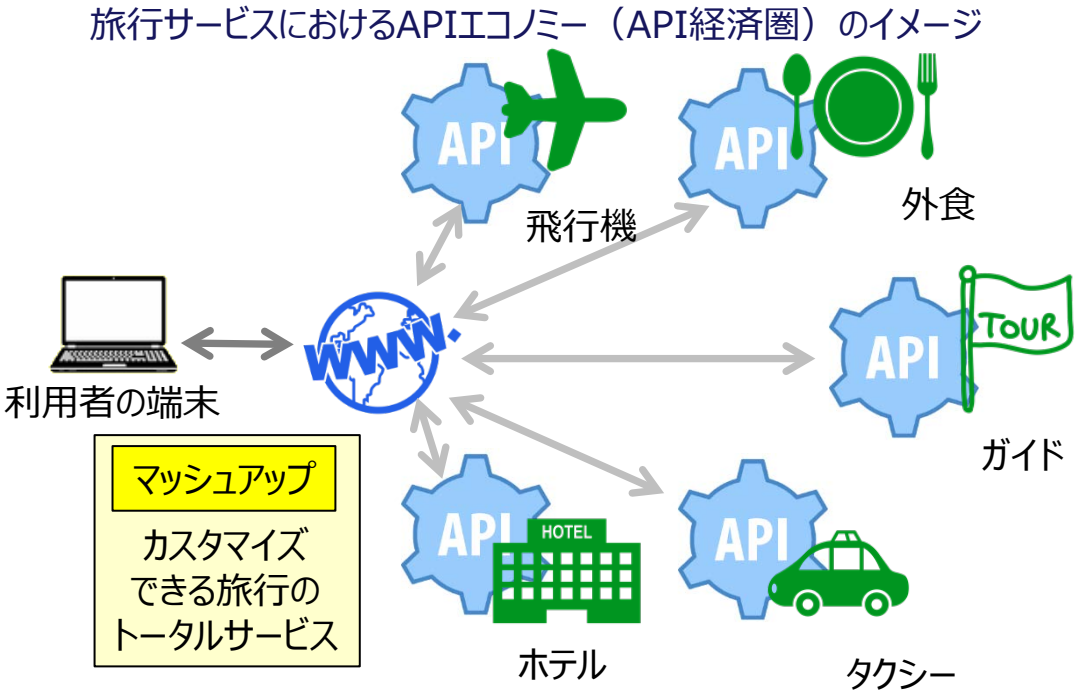
APIエコノミー（API経済圏）

◆オープンAPIが普及することによって、ビジネスをつなぐAPIエコノミーが着目されています。

- APIをビジネスに活用すると、自社の得意分野や新規事業に注力することができます。
  - 「食べログ」「ぐるなび」「ホットペッパーグルメ」は、地図表示に「Googleマップ」のAPIを活用し、飲食店の情報表示や紹介に注力しています。
- APIをビジネスに活用することやAPIを公開・利用するエコノミー（経済圏）を**APIエコノミー**と呼びます。
  - 技術的なAPIは「コンピュータ（プログラム）」をつなぐ役割に着目しますが、「APIエコノミー」と言った場合は「ビジネスをつなぐ役割」に着目します。
  - APIによって様々なビジネスの組み合わせを行うことが可能になり、例えば、「カスタマイズできる旅行のトータルサービス」に活用できます。
- 2015年におけるIBMの予測では、APIエコノミーの市場規模は2018年までに2兆2000億ドルになるとされました。

【出所】IBM Unveils Matchmaking Technology to Navigate API Economy［IBM］
https://www-03.ibm.com/press/us/en/pressrelease/48026.wss





オープンAPIの利点とIoTとの親和性

◆オープンAPIの利点は、IoTとAPIの親和性につながっています。

APIの利点

（1）情報提供側が更新した情報はAPIに反映されるため、ウェブサイトやアプリにおける情報のアップデートは不要で、常に最新の情報を表示できます。

（2）APIを利用すると、公的統計e-Statの全情報、Googleマップの全世界の地図といった膨大な情報にアクセスできるとともに関心のあるデータ・自地域の周辺といった対象を絞った情報にアクセスできます。

（3）APIでは音声や画像を送受信でき、人工知能による分析やサービスなど、最新技術の活用が可能です。

（4）様々なAPIサービスの情報をマッシュアップによって組み合わせて、新たな複合サービスを作成できます。

IoTとの親和性

（1）（2）より、最新かつ膨大な情報の中から対象とする情報を絞ってアクセス可能であり、記憶容量、通信速度の制限が強いIoT機器でも利用することができます。

（3）より、高負荷の処理はAPIを通じて送信先のクラウド等で行えるため、処理能力の制限が強いIoT機器でも、最新技術が利用できます。

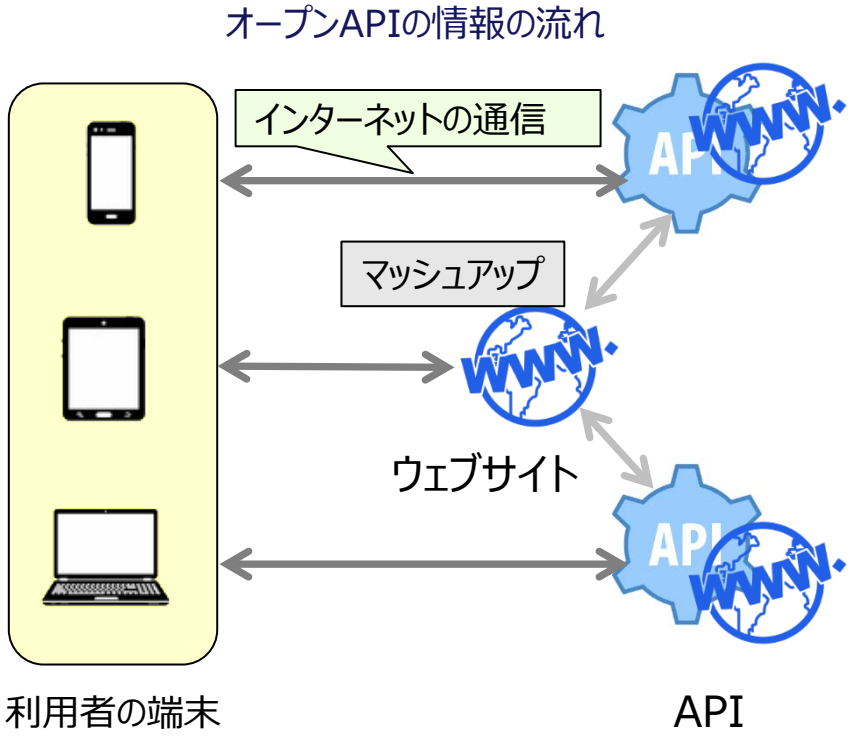
（4）より、APIは複数の情報ソースから表示や内容をカスタマイズした情報提供ができるため、スマートフォン用の複合情報の表示など、個々のIoT機器に向いた情報提供が可能です。

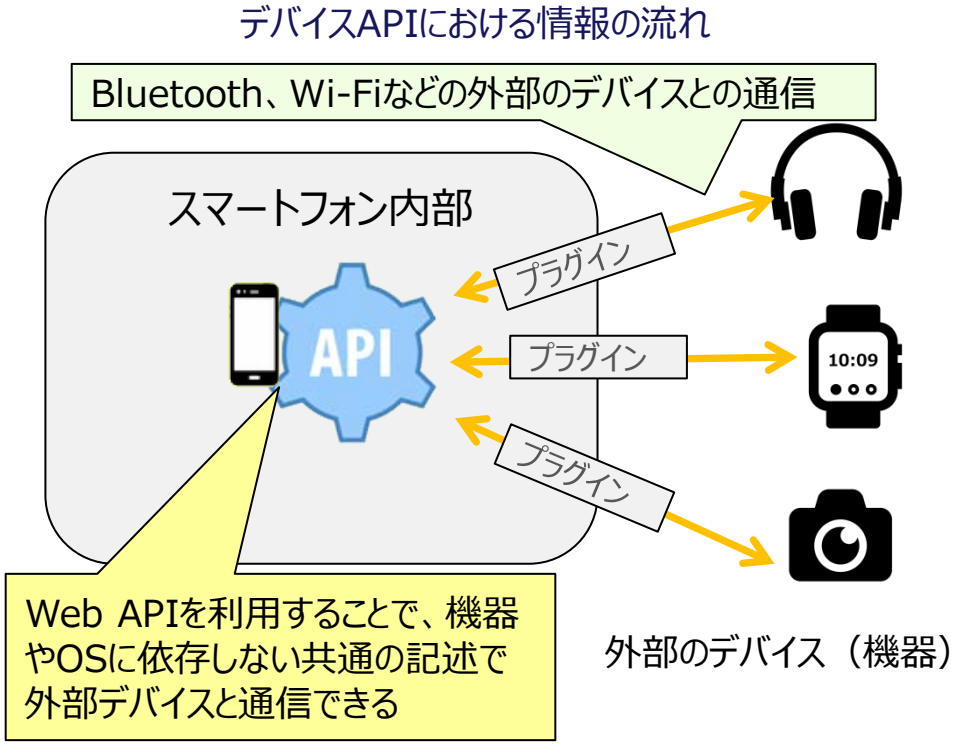
- 講座4-3【参考2】では、Rのプログラミングでe-Stat APIを利用する手順を説明し、(1)(2)の利点を活かす形で最新の情報かつ膨大な情報の一部にアクセスする事例を具体的に示しています。
- オープンAPIの利点は、IoT機器における使いやすさにつながっています。一方で、スマートフォンをはじめとするIoT機器との親和性が高い「デバイスWeb API」という技術もあります。

デバイスWeb API

◆デバイスWeb APIは、一般にスマートフォンとデバイスの情報の送受信を仲介します。

- APIの技術の中には、スマートフォン等にインストールされ、ヘッドフォンやスマートウォッチといった外部のデバイス（機器）との通信に利用される**デバイスWeb API**があります。
  - 本教材のこれまでのスライドで例示したオープンAPIは、無関係の外部組織等からインターネット通信でAPIを利用する形式でした。
- デバイスWeb APIは、スマートフォンと外部のデバイスの通信を汎用的なHTTP形式で仲介します。
  - オープンAPIとデバイスWeb APIは、標準的なウェブサイト閲覧の通信方法であるHTTP通信を使うことで、機器やOSに依存しない情報の送受信を可能としている点は共通しています。



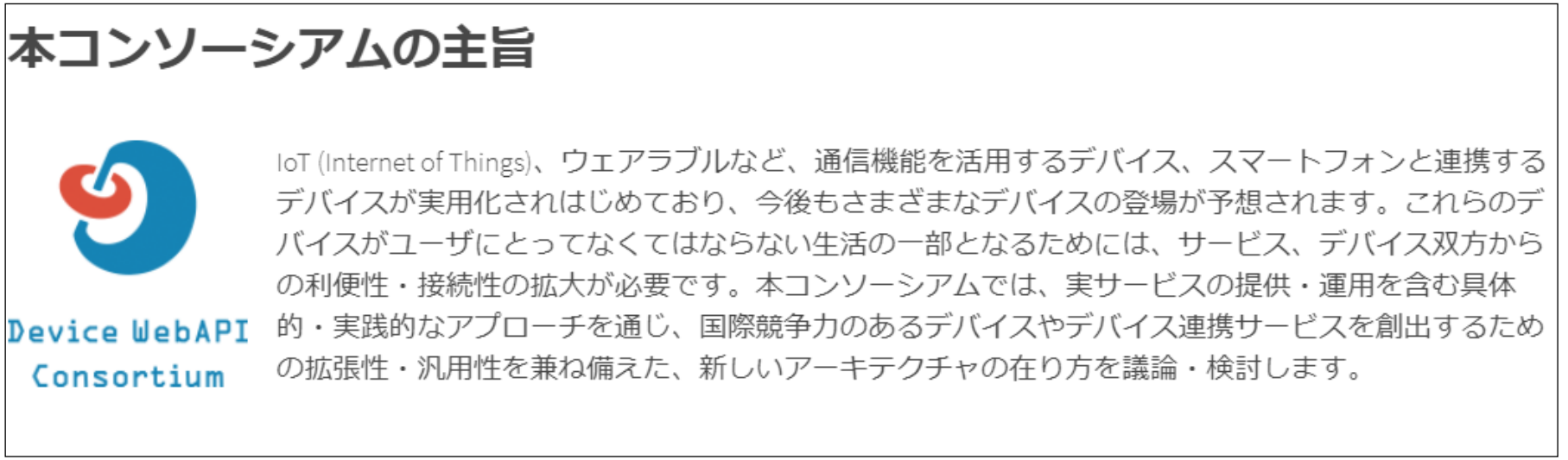


デバイスWeb API コンソーシアム

◆デバイスWeb APIの普及のため、「デバイスWeb API コンソーシアム」が設立されました。

- 共通の「デバイスWeb APIをスマートフォンにインストールすれば、外部のデバイスを制御するための情報の形式がスマートフォンの機種やOSに依存せず、統一されます。
  - 統一された情報の形式に合わせて、外部のデバイスを制御するアプリ（プラグイン）を作成する場合は、アプリの開発負担が軽減されます。
- 共通の「デバイスWeb API」を開発・策定するべく**デバイスWeb APIコンソーシアム**が2015年4月に設立され、デバイスWeb APIにおける規格「GotAPI」の普及や機能検討を行っています。
  - 「デバイスWeb APIコンソーシアム」は、株式会社NTTドコモとソフトバンク株式会社によって設立されました。
  - 2018年3月において、「デバイスWeb APIコンソーシアム」は、KDDI株式会社を含め、120の企業会員で構成されています。

デバイスWeb APIコンソーシアムの趣旨



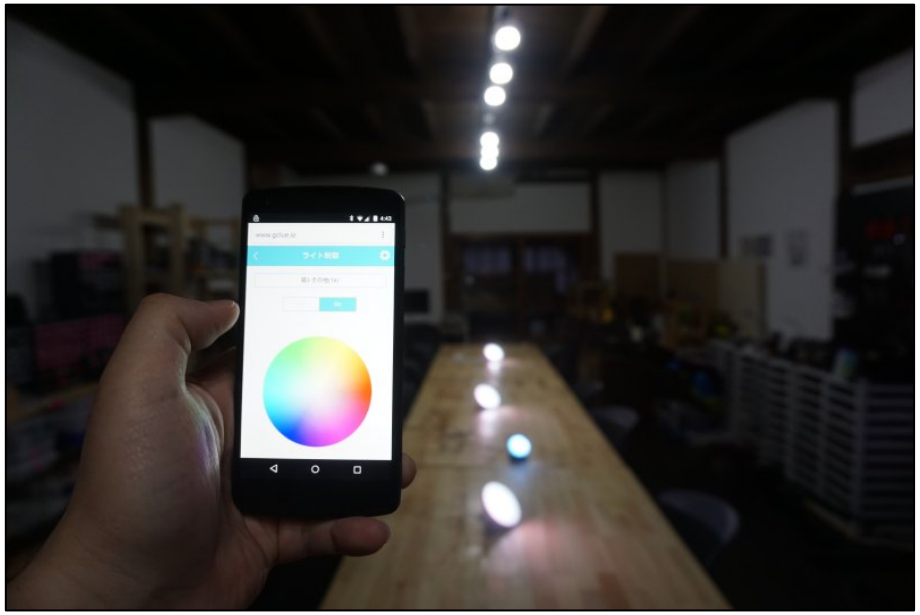
【出所】デバイスWeb APIコンソーシアム　https://device-Web API.org/

デバイスWeb APIの事例

◆デバイスWeb APIによって、スマートフォンから様々なIoT機器を操作できるようになります。

- デバイスWeb APIのスマートフォンアプリ「Device Web API Manager」を利用すれば、様々なIoT機器を操作することができます。
  - 操作可能なIoT機器として、Philips 社が販売しているIoT電灯のhue (ヒュー)や株式会社リコーが販売している360°カメラのTHETA（シータ）が挙げられます。

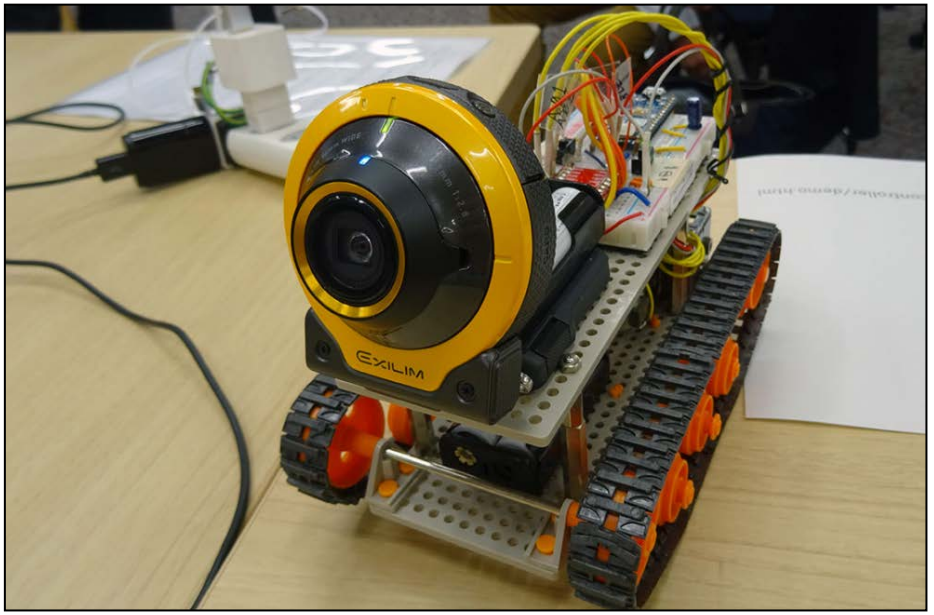
Device Web API Managerによって電灯を操作する様子



【出所】Device Web API Manager
http://www.gclue.io/dwa/

- カシオ計算機株式会社では、デジタルカメラEXILIM（エクシリム）の一部の機種をデバイスWeb APIに対応させ、WebAPI のプラグインを一般公開しています。
  - 2017年4月に開催されたデバイスWeb APIコンソーシアムの第三回総会では、カメラを搭載した戦車をスマートフォンからデバイスWeb APIを通して操作するデモを行いました。

デバイスWeb APIによって制御されるカメラ搭載戦車



【出所】第三回総会レポート［デバイスWeb APIコンソーシアム］
https://device-Web API.org/event/20170413/3rd_plenary_meeting.html

人間にとって、読みやすく使いやすいHTML

◆ウェブブラウザで利用するHTMLは、人間にとって見やすく、使いやすく作られています。

- このパートでは、Web APIにおいて利用頻度が高いファイル形式（**XML**、**JSON**）を**HTML**との対比で紹介します。
- ウェブサイトの表示には、**HTML（エイチティーエムエル）というファイル形式**が一般的に利用されています。
  - HTMLは、「HyperText　Markup Language」の略で、HyperText とはウェブページのリンクが組み込めることを指しています。
- HTMLでは、レイアウトを工夫したり、画像を配置したりすることで、人間にとって理解しやすい表示ができます。
  - HTMLでは強調したい語句は、大きなフォントを使ったり、太字にしたりすることで、人間の注目を集めるように表現することができます。
- HTMLをウェブブラウザで表示すると、太字の強調表示や画像挿入があり、人間にとって分かりやすくなっています。
  - PC利用とスマートフォン利用に応じて、表示方法を変更することも可能ですが、閲覧しているHTMLファイルは共通しています。

Wikipediaにおける「World Wide Web」の記事の表示（左側：PC　右側：スマートフォン）

【出所】Wikipedia　https://ja.wikipedia.org/wiki/World_Wide_Web

HTMLの主な目的

◆HTMLの主な目的は、人間にとって読みやすく、使いやすい表示をすることです。

- Webサイトで一般的なHTMLでは、大きめの文字や太字が人間にとって、自然と目につく強調箇所となります。

HTMLファイルをテキストエディターで開いた内容（ソース）

```
<!DOCTYPE html>
<html>
<head>
<title>HTMLの例示</title>
</head>
<body>
<h1>大見出し</h1>
<h2>中見出し</h2>
<p>文字を<b>太字</b>や
<font color="red">赤字</font>で書ける。</p>
<a href="http://www.soumu.go.jp/">総務省</a>へのリンク
<p>↓のように画像も入れることができる。</p>
<img src="spade.png" title="スペード">
<img src="heart.png" title="ハート">
</body>
</html>
```

Windowsの「メモ帳」などのテキストエディターでHTMLファイルを開いた場合の表示

コンピュータ（プログラム）は、HTMLの見せ方を解釈してWebブラウザに表示します。
（ただし、コンピュータ（プログラム）にとって、ファイルの内容は把握しにくい）

ウェブブラウザによるHTMLの表示

Internet Explorer、Chrome、Safariといったウェブブラウザで HTMLファイルを開いた場合の表示

ウェブブラウザの表示は、フォントの大きさや文字装飾から、人間にとって強調点が把握しやすくなっています。人間は画像情報も理解できます。

- HTMLでは、クリックするとウェブサイトを移動できるリンクを表示でき、人間にとって分かりやすいだけではなく、実用的に使いやすいファイルとなっています。

コンピュータ（プログラム）が内容を読みやすいファイル形式

◆コンピュータ（プログラム）が、内容を読みやすい形式としてXMLやJSONが挙げられます。

- コンピュータ（プログラム）にとっては、内容を把握しやすいファイルは、情報の構造・意味が明確なファイルです。
  - 例えば、括弧開きと括弧閉じが正確に対応しており、全体の情報における各情報の位置によって情報の意味が分かれば、情報の構造は明確と言えます。
  - 人間が読む文章においても、『鍵括弧「」の中は会話文、丸括弧（）の中は読み方や補足説明』といった括弧に基づく情報構造やルールがあります。
- 情報の構造が、明確なファイル形式として、**XML（エックスエムエル）**や**JSON（ジェイソン）**が挙げられます。
  - XMLは「Extensible Markup Language」、JSONは「JavaScript Object Notation」の略です。
  - XMLやJSONは、講座2-1および講座3-1で示すデータの分類において半構造化データに該当する形式です。
- XMLやJSONはコンピュータ（プログラム）にとって内容を把握しやすい一方で、人間にとっては読みにくい形式です。
  - スライド下部の左側にはWikipediaの「World Wide Web」のXMLでの出力、スライド下部の右側にはJSONでの出力を示しています。

「World Wide Web」に関するXML出力


```
▼<api batchcomplete="">
  ▼<query>
    ▼<normalized>
       <n from="World_Wide_Web" to="World Wide Web"/>
     </normalized>
    ▼<pages>
      ▼<page _idx="2459" pageid="2459" ns="0" title="World Wide Web">
        ▼<revisions>
          ▼<rev contentformat="text/x-wiki" contentmodel="wikitext" xml:space="preserve">
              {{Redirect2|WWW|ウェブ||ウェッブ}} {{Otheruses||世界初の[[ウェブブラウザ]]〈閲覧
              ターネット}}、{{lang|en|'''World Wide Web'''}}（'''ワールド・ワイド・ウェブ'''、
              {{lang|en|'''Web'''}}、'''ウェブ'''、{{lang|en|'''W3'''}}（ダブリュー スリー）<r
              俗には「インターネット」という表現がワールド・ワイド・ウェブを指す場合もある。
              World Wide Web をグラフィック表現したもの。[[複雑ネットワーク]]の一例でもある。]
              Markup Language|HTML]]や[[Extensible HyperText Markup Language|XHTML]]といった[[
              れるハイパーテキストとは、文書中に別の文書の[[Uniform Resource Identifier|URL]]
              士を相互に参照可能にするシステムである。閲覧者は表示している文書中でハイパーリン
```


【出所】XML format [Wikipedia]
https://ja.wikipedia.org/w/api.php?format=xml&action=query&prop=revisions&rvprop=content&titles=World_Wide_Web

「World Wide Web」に関するJSON出力

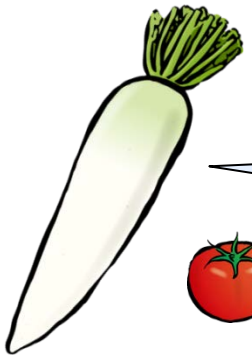

```
{
    "batchcomplete": "",
    "query": {
        "normalized": [
            {
                "from": "World_Wide_Web",
                "to": "World Wide Web"
            }
        ],
        "pages": {
            "2459": {
                "pageid": 2459,
                "ns": 0,
                "title": "World Wide Web",
                "revisions": [
                    {
                        "contentformat": "text/x-wiki",
                        "contentmodel": "wikitext",
                        "*": "
{{Redirect2|WWW|¥u30a6¥u30a7¥u30d6||¥u30a6¥u30a7¥u30c3¥u30d6}}¥n{{Otheruses||¥u4e16¥u75
(¥u7b11)}}¥n¥n{{¥u30a4¥u30f3¥u30bf¥u30fc¥u30cd¥u30c3¥u30c8}}¥n{{lang|en|'''World Wide
```


【出所】JSON format [Wikipedia]
https://ja.wikipedia.org/w/api.php?action=query&prop=revisions&rvprop=content&titles=World_Wide_Web

XMLやJSONの主な目的（Web APIにおける利用）

◆XMLやJSONの主な目的は、コンピュータ（プログラム） 同士で情報を伝達することです。

- XMLやJSONは、コンピュータ（プログラム）にとって、内容が分かりやすい構造で情報が格納されています。
  - 数値データに関しては、人間にとって分かりやすい表示として二次元の表形式が挙げられます。（HTML内で表を描くことも可能です。）
- XML、JSONでは、商品名をname、価格をpriceといった形で利用者独自の**記号（マークアップ）**が使えます。
  - HTMLにおける記号（マークアップ）は、<h1>が大見出し、<b>が太字といった形で、W3C（World Wide Web Consortium）という標準化団体があらかじめ定めた記号（マークアップ）のみ利用できます。［W3CによるHTMLの仕様：https://www.w3.org/html/］
- XMLやJSONは、Web APIにおける利用など、コンピュータ（プログラム）同士の情報共有に適したファイル形式です。
  - 情報共有を行う者同士で、記号（マークアップ）のルールをあらかじめ決めておくことで、円滑な情報共有ができます。
  - 一般にJSONの方がXMLよりも簡潔であるため、新たに取り決められるWeb APIの情報交換のルールにおいては、JSONの利用が主流となっています。



野菜：大根
120円

野菜：トマト
80円

人間にとって読みやすい

表形式

id	kind	name	price
1001	野菜	大根	120
1002	野菜	トマト	80

<XML>

XMLの記述

```
<goods>
 <item>
  <id>1001</id>
  <kind>野菜</kind>
  <name>大根</name>
  <price>120</price>
 </item>
 <item>
  <id>1002</id>
  <kind>野菜</kind>
  <name>トマト</name>
  <price>80</price>
 </item>
<goods>
```

{JSON}

JSONの記述

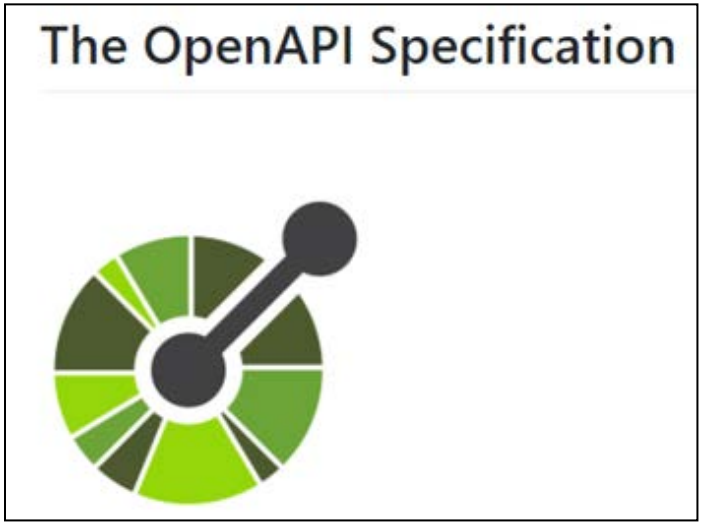
```
{"goods":[
            {
            "id":"1001",
            "kind":"野菜",
            "name":"大根",
            "price":"120"
            },
            {
            "id":"1002",
            "kind":"野菜",
            "name":"トマト",
            "price":"80"
            },
]}
```

API用データの記述の標準化

◆JSON等のWeb API用データは、その記述の標準化が進められています。

- 2015年11月オープンAPIの記述標準化を目的とする**「Open API Initiative（オープン エーピーアイ イニシアチブ）」**が結成されました。
  - Open API Initiativeの結成時から参加している企業として、Microsoft、Google、IBMが挙げられます。
- Open API Initiativeから公開されているAPIの記述作成ツールを利用すれば、標準化されたルールに沿ったAPI用データ（JSON等）の記述の設計を容易に行うことができます。

Open API Initiativeのロゴ



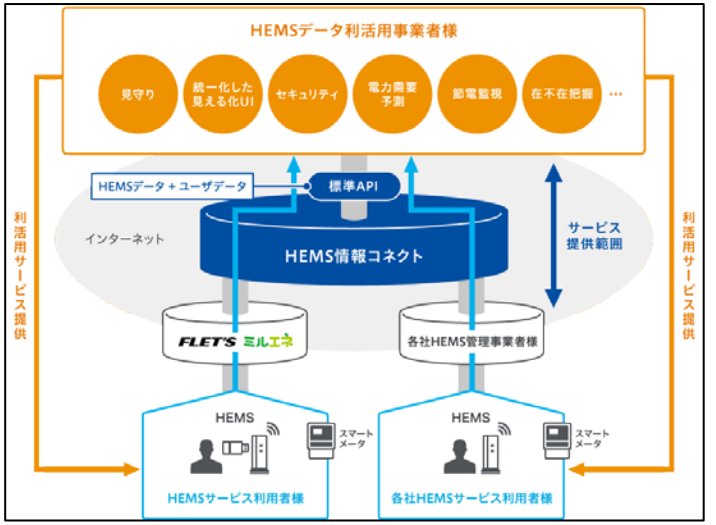
【出所】Open API Initiative
https://www.openapis.org/

- 経済産業省では、2016年3月に家屋の電力制御を行うHEMSのAPI用データ（XML、JSON）に関する標準化に関する仕様書をとりまとめました。

【出所】HEMSデータ利活用事業者間 API標準仕様書（1.0版）[経済産業省]
http://www.meti.go.jp/committee/kenkyukai/shoujo/smart_house/pdf/009_s10_00.pdf

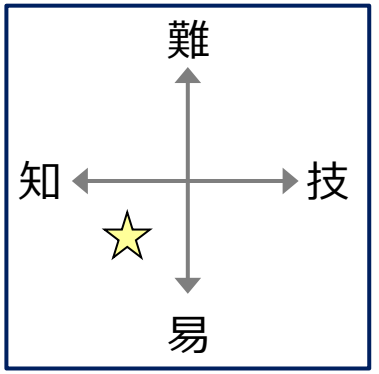
- NTT東日本では、既存のHEMSが出力するデータを変換して標準化された統合データを提供する「HEMS情報コネクト」を2016年10月に開始しました。

HEMS情報コネクトのサービスイメージ



【出所】HEMS情報コネクト［NTT東日本］
https://flets.com/hemsconnect/service.html

総務省　ICTスキル総合習得教材



［コース2］データ蓄積

2-1：データ蓄積用サーバとしてのクラウド

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_2_1.pdf

	1	2	3	4	5
［コース1］データ収集					
［コース2］データ蓄積	●				
［コース3］データ分析					
［コース4］データ利活用					

本講座の学習内容［2-1：データ蓄積用サーバとしてのクラウド］

【講座概要】

- データベースとサーバの概要と役割を説明します。
- リレーショナルデータベースを中心にデータベースマネージメントシステムの概要を紹介します。
- サーバーの仮想化とスケーラビリティー（拡張性）の関係を説明します。
- クラウドの出自やアメリカの国立標準技術研究所が挙げた5つの特徴、クラウド以外のサーバ運用方法との関係を説明します。

【講座構成】

座学

- [1] データベースとサーバ
- [2] データベースの概要
- [3] サーバの概要
- [4] サーバの仮想化
- [5] クラウドの概念

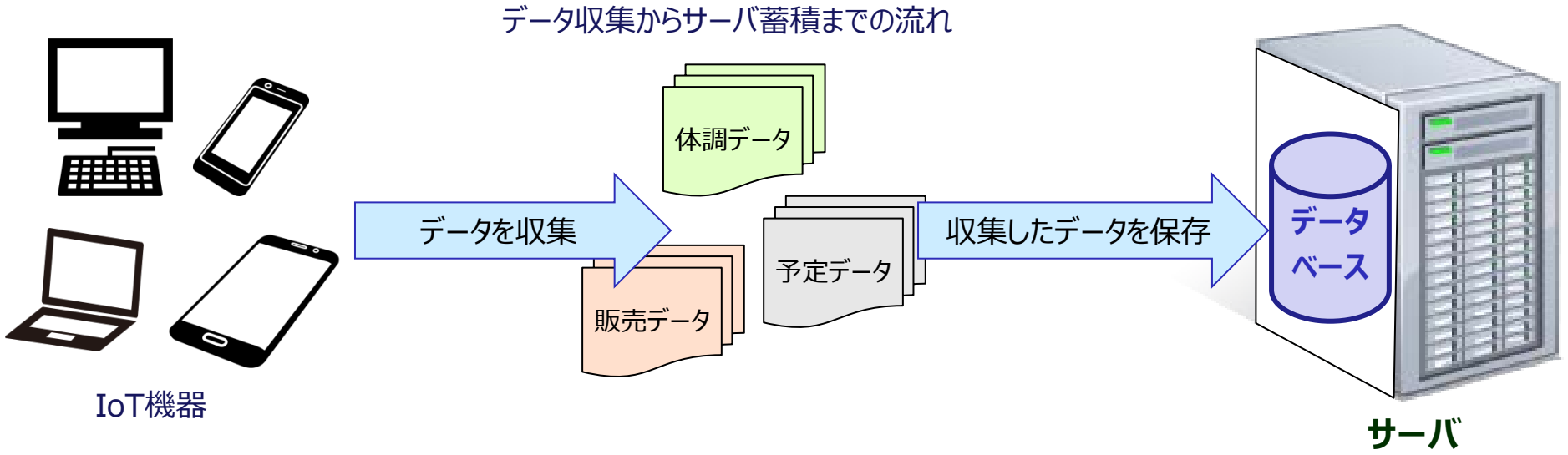
【学習のゴール】

- ✓ データベースとサーバの概要を把握する。
- ✓ データベースマネージメントシステムの概要とリレーショナルデータベースの形式を把握する。
- ✓ サーバの仮想化とスケーラビリティー（拡張性）を理解する。
- ✓ 「クラウド」の出自と5つの特徴を紹介できる。

インターネット上のデータベースサーバ

◆様々なIoT機器から得られたデータは、データベースサーバに集約して蓄積できます。

- IoT機器等で収集したデータは、インターネットに接続された サーバ 上の**データベース （⇒データベースサーバ）**に蓄積することで長期的な保存をすることが一般的です。
- データを**データベース**へ保存することで、データの整理や抽出が容易になります。
- データの保存先として、**サーバ**を利用することで、複数台、複数種のIoT機器から得られたデータを集約して蓄積することができます。
  - サーバおよびデータベースに関しては続くスライドで説明します。



□ この講座では、データベースサーバを構成するサーバとデータベースをそれぞれ説明した後にサーバの運用形態の一つである「クラウド」を説明します。

データベースの意味

◆検索が容易となるルールに従って整理されたデータの集まりをデータベースといいます。

- **「検索が容易となるルール」**に従って、整理されたデータの集まりを**データベース**といいます。
- 著作権法（十の三）において、データベースは「論文、数値、図形その他の情報の集合物であって、それらの情報を電子計算機を用いて検索することができるように体系的に構成したものをいう。」と記述されています。
- 検索が容易となるルールに従って整理された名簿、電話帳データはデータベースです。
  - 情報通信に関する資料において、データベースは、円柱 または円柱が連なった形 での表記が、通例となっています。

データベースの例（従業員名簿）

氏名	社員番号	部門名	所在地	電話番号
岡部翔	1234	営業部	東京	03-1234-****
小倉大輔	5678	営業部	名古屋	052-2345-****
黒川聡	9012	経理部	東京	03-1234-****
高山亮	3456	人事部	東京	03-1234-****
星野大介	7890	人事部	大阪	06-3456-****
山岸直樹	9876	開発部	沖縄	045-4567-****

- テキスト形式などで同じ情報量であったとしても、検索が容易となるルールに従って整理がされていなければ、データベースとは言えません。

サーバの意味

◆「サーバ」は「供給をするもの」という意味があり、サーバコンピュータはデータを供給します。

- **サーバ（server）**には、要求に応じて**「供給をするもの」**という意味があります。
- インターネット上にあるサーバコンピュータは、いったんデータを溜め込んで要求に応じて「データを供給するもの」です。
  - サーバコンピュータは略称として「サーバ」と呼ぶことが定着しています。
  - 家庭や飲食店にある「ウォーターサーバ」は、いったん水を溜めて要求に応じて水を「供給するもの」であり、ケーキショップにある「ケーキサーバ」は、いったんケーキを上に乗せてから要求に応じてケーキを「供給するもの」です。
- サーバという名称は、要求に応じての「供給」に着目していますが、供給する前には「蓄積」する必要があります。
  - サーバコンピュータは、データを供給する前にデータを蓄積します。

サーバコンピュータ

ウォーターサーバ

ケーキサーバ

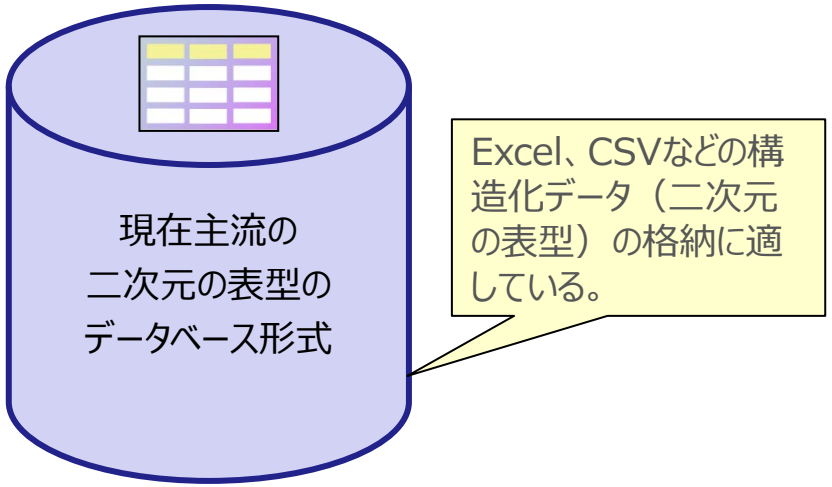
データベースの種類

◆データベースは、「リレーショナルデータベース」と「NoSQLデータベース」に大別できます。

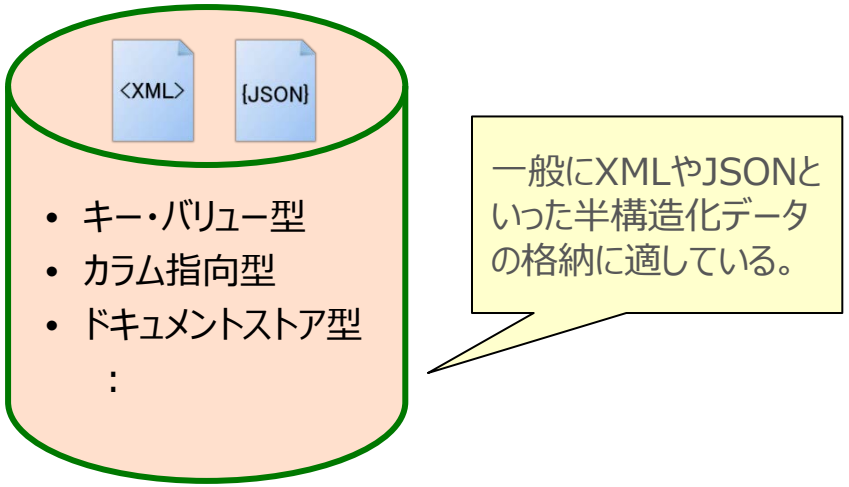
- 2017年現在、**リレーショナルデータベース（関係データベース）**が、主流のデータベース形式です。
  - リレーショナルデータベースは、ExcelファイルやCSVなどの二次元の表型のデータの格納に適しています。
  - リレーショナルデータベースは情報通信用語として、英語表記の「Relational DataBase」から「RDB」と略すこともあります。

- **NoSQLデータベース**は「Not only SQL」の略とされ、リレーショナルデータベースの短所の補完、ビッグデータの活用に関連して現れたデータベースの構造の総称で、近年着目されているデータベースの形式です。
  - NoSQLの中にも「キー・バリュー型」「カラム指向型」「ドキュメント指向型」などのデータベースの構造の分類があります。
  - NoSQLデータベースには、複数のコンピュータにまたがる保存に適したものやXMLやJSONといった半構造化データの格納に適するものがあります。
  - XMLやJSONは、講座1-5に示したAPIによる情報伝達で一般的なファイル形式でもあります。

リレーショナルデータベース型



総称としてのNoSQLデータベース



構造化データ、半構造化データ、非構造化データ

◆データは、その構造によって3種類に分類できます。

- データはその構造によって**構造化データ**、**半構造化データ**、**非構造化データ**に大別されます。
  - 講座3-1のビッグデータの特性においても、「構造化データ」「半構造化データ」「非構造化データ」を紹介します。
  - 半構造化データの「XML」「JSON」に関しては、講座1-4のAPIで利用されるファイル形式として説明しました。

「構造化データ」「半構造化データ」「非構造化データ」の説明

データ種別	説明	データ形式の例
構造化データ	二次元の表形式になっているか、データの一部を見ただけで二次元の表形式への変換可能性、変換方法が分かるデータ	CSV、固定長、Excel（リレーショナルデータベース型）
半構造化データ	データ内に規則性に関する区切りはあるものの、データの一部を見ただけでは二次元の表形式への変換可能性・変換方法が分からないデータ	XML、JSON
非構造化データ	データ内に規則性に関する区切りがなく、データ（の一部）を見ただけで、二次元の表形式に変換できないことが分かるデータ	規則性に関する区切りのないテキスト、PDF、音声、画像、動画

二次元の表形式の構造化データ

世帯名	大人1	大人2	子供1
山田家	世帯主	妻	長女

XML形式の半構造化データ

```
<世帯>
  <世帯名>山田家</世帯名>
  <大人>世帯主</大人>
  <大人>妻</大人>
  <子供>長女</子供>
</世帯>
```

画像形式の非構造化データ

- 一般に構造化データはリレーショナルデータベースの形式、半構造化データはNoSQLの形式が格納に向いています。

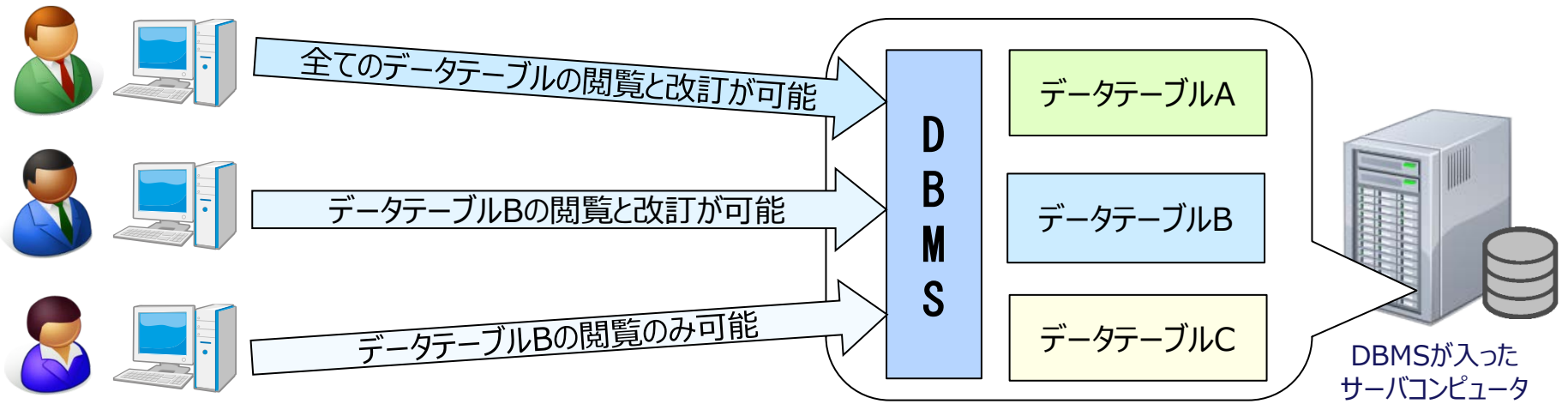
データベース管理システム（DBMS）

◆データベース管理システム（DBMS）と呼ばれるデータベース管理のソフトウェアがあります。

- データベースを管理するためのソフトウェアを**データベース管理システム（DBMS: database management system）**といいます。

- DBMSを利用すると、複数の利用者での利用する場合でも、データベース単位、データテーブル単位、データレコード単位で、閲覧や改訂の権限を分けることができるなど、データの安全管理が容易になります。
  - データベースとExcelファイルでの対応では、データベースはファイル全体、データテーブルは各シート、データレコードはシート内の各行のイメージです。

複数の利用者でのDBMSの利用イメージ

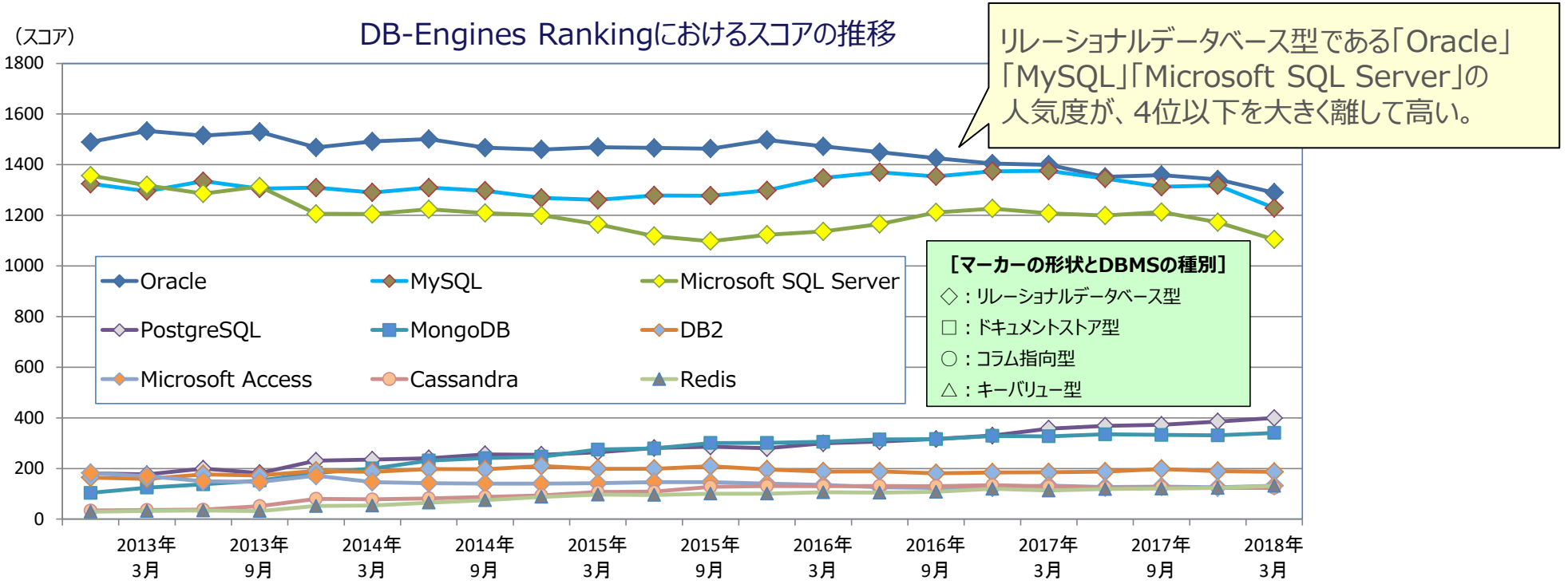


- 代表的な商用DBMSとして、「Oracle （Database）」「Microsoft SQL Server」が挙げられ、一定の条件下において無償利用ができるDBMSとして「MySQL」「PostgreSQL」が挙げられます。
  - これら4種の代表的なデータベース管理システムは、いずれも「リレーショナルデータベース型」と呼ばれる構造化データの格納に適した現在主流の形式です。
  - Microsoft Office Professionalを構成する一般のPC用ソフトウェアの「Microsoft Access」も、リレーショナルデータベース型のDBMSです。

DBMS（データベースマネジメントシステム）の人気ランキング

◆DBMSの人気ランキングサイトによると、上位3つのDBMSに人気が集中しています。

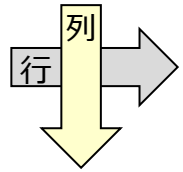
- DB-Engines Rankingでは、各DBMSの人気を点数化して、2012年からの推移を月単位で公開しています。
  - このランキングにおいては「関連ウェブサイトの数」「検索頻度」「ウェブ掲示板での言及数」「求職・求人サイトで言及数」「SNSでの言及数」によって、世界規模で各DBMSの人気度を測定しています。
- このランキングによると、2018年3月時点において「Oracle」「MySQL」「Microsoft SQL Server」の3つのDBMSが、他のDBMSを大きく離して人気が高いことが示されています。
  - 下記グラフは2018年3月時点における人気上位9つのDBMSの推移を描いています。「MongoDB」がドキュメントストア型、「Cassandra」がコラム指向型、「Redis」がキーバリュー型であり、それ以外の6つのDBMSは「リレーショナルデータベース型」です



【出所】DB-Engines Ranking［solid IT］の情報に基づき作成 https://db-engines.com/en/ranking

リレーショナルデータベースの構造

◆リレーショナルデータベースでは、2次元の表形式でデータを表します。



- リレーショナルデータベースでは、2次元の表形式でデータセットを表します。
  - リレーショナルデータベースでは、表形式で表されたデータの集まりをそのまま表（table）と呼びます。
- リレーショナルデータベースにおいては、水平線側を**行（row）**、垂直線側を**列（column）**と呼びます。
  - 漢数字の「十」の書き順で「行・列」と覚えてください。
- 水平側の行には個別の人、物、事象の情報が記入され、垂直側の列にはその人、物、事象の属性が記入されます。

- 名簿のデータベースでは、行には個別の情報が記入され、列には「氏名」「社員番号」「部門名」といった属性が記入されます
- リレーショナルデータベースでは、社員番号のような重複する可能性のない項目を、主キー（primary key）として設定します。
- リレーショナルデータベースでは、部門番号で「従業員リストと部門リストをつなぐ」ように、複数の表をつないで利用可能です。

従業員名簿でのリレーショナルデータベースの例

従業員リスト

	1列目	2列目	3列目
列名⇒	**社員番号**	**氏名**	**部門番号**
1行目	1234	岡部翔	10
2行目	2345	小倉大輔	10
3行目	3456	黒川聡	20
4行目	4568	高山亮	30
5行目	5678	星野大介	30
6行目	6789	山岸直樹	40

部門リスト

部門番号	**部門名**	**所在地**
10	首都圏営業部	東京
20	首都圏経理部	神奈川
30	関西人事部	大阪
40	九州開発部	福岡

SQL（Structured Query Language）

◆SQLは、シンプルな英単語でリレーショナルデータベースを操作できるコンピュータ言語です。

- **SQL（Structured Query Language）**は、リレーショナルデータベースを操作するためのコンピュータ言語です。
  - リレーショナルデータベース型以外のDBMSのNoSQLは、一般に「Not only SQL」の略であるとされており、SQLを使わないのではなく、「SQLだけではない」「SQL言語を使わなくてもデータの操作ができる」という解釈になっています。
- SQLはシンプルな英単語による記述でデータベースを操作することができます。
  - 例えば、「住所録（表）から氏名（列）が『佐藤』で始まる氏名リストを表示する」場合は、次のようにSQLのコマンドを記述します。

氏名が『佐藤』で始まる氏名リストを表示するSQLコマンド

```
select 氏名
from   住所録
where  氏名 like '佐藤%';
```

SQLにおいて、%（パーセント）は無記入を含む任意の文字列を表し、この記述で氏名が「佐藤」で始まる全ての氏名を表示します。

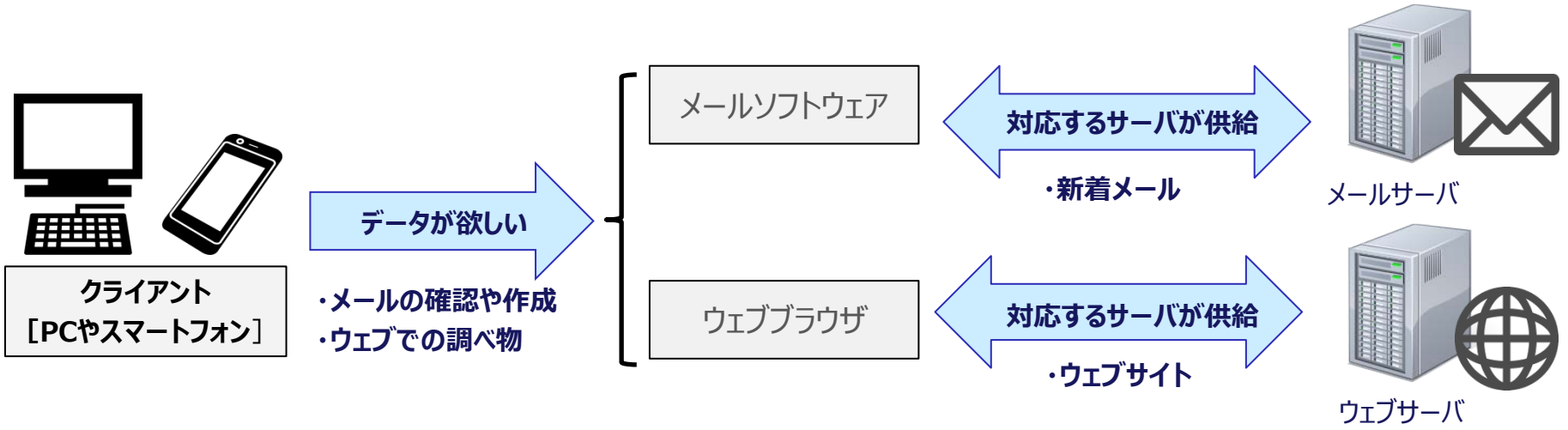
- SQLを使うと、**トランザクション**と呼ばれる一括したデータベースへの指示をすることもできます。
  - トランザクションを使って、「19時の予約をキャンセルして、20時の予約をする」とデータベースに指示すると、20時に予約できない場合には19時の予約がキャンセルされません。
- 「Microsoft Office Professional」を構成するソフトウェアの「Microsoft Access」では、SQLによるデータベースの操作が可能です。
  - Microsoft Excelにおいてもデータの整理方法に配慮すれば、リレーショナルデータベース形式のデータベースを作ることができます。一方、Microsoft Accessでは数値を入れるべき列には数値しか入力できないなど、リレーショナルデータベース型であることが強制されています。

インターネット上のサーバの種類

◆インターネット上には、様々な種類のサーバがあり、要求に応じて各データを提供します。

- **サーバコンピュータ**は、利用権限のある様々な機器からの要求に応じて、データを蓄積・供給します。
  - PCとスマートフォンの異なる機器からでも同一のサーバコンピュータが利用でき、複数のユーザーで同一のサーバを利用することもできます。
- インターネット上のウェブサーバは、PCやスマートフォンのウェブブラウザからの要求に応じて、HTMLをはじめとするウェブサイトのデータを提供します。
- サーバに要求を出す機器は**クライアント**と呼ばれ、ウェブブラウザを搭載したPCやスマートフォンは、ウェブサーバに対するクライアントです。
  - クライアントは「依頼者」「要求をするもの」の意味し、情報通信の用語において、サーバ「供給するもの」の対義語となっています。

クライアントの要求に応じたサーバの仕組み



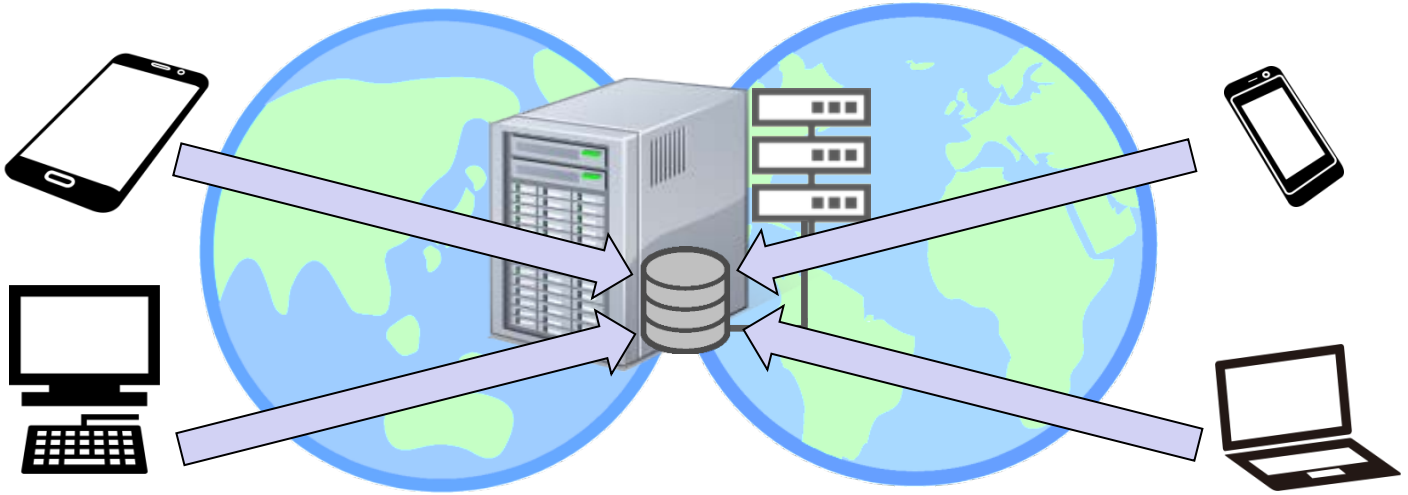
□ インターネット上にあるメールサーバ、ウェブサーバは、世界のどこからでもアクセス可能です。

インターネットからアクセスできるデータベースサーバ

◆DBMSをインターネットからアクセス可能とすることで、世界中からデータを利用できます。

- DBMS（データベースマネージメントシステム）が搭載されたサーバを**データベースサーバ**と呼び、インターネットを経由したアクセスを可能とするケースがあります。
- データベースサーバがインターネット上からアクセスできれば、メールサーバやウェブサーバと同様に利用権限のある世界中の利用者、機器から利用できます。
  - 複数のIoT機器から得られた情報を集約して蓄積でき、複数のユーザーがデータを閲覧・更新することが可能です。
- データベースサーバではファイルの保存場所を記録することで、ネットワーク上でのファイル管理にも利用されています。
  - データベースサーバとファイルの保存場所（ファイルサーバ、オンラインストレージ）は一般に異なりますが、データベースサーバには検索しやすくファイルの保存場所・利用権限を記録しておくことで、ネットワーク上での円滑なファイル管理、ファイル利用にも貢献しています。

世界中からアクセスできるインターネット上のデータベースサーバのイメージ

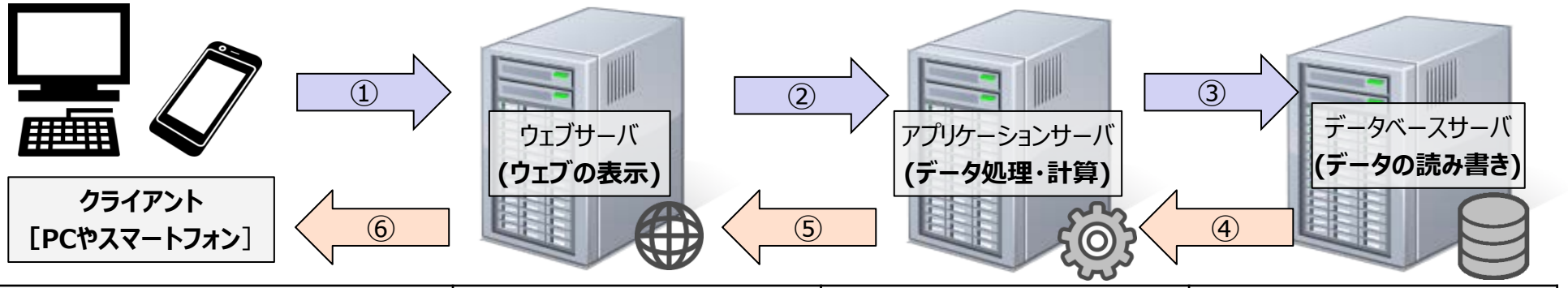


ウェブサービスの背後にあるデータベースサーバ

◆様々なウェブサービスは、背後でデータベースサーバが稼働することによって成り立っています。

- ウェブサービスにおけるデータベースサーバは、ウェブサーバやアプリケーションサーバの背後にあり、**ウェブ3層**と呼ばれる構造をとっていることが一般的です。
  - ウェブ3層と言われるウェブサービスの構造では、ウェブサーバはウェブサイトの表示・受付を担い、アプリケーションサーバはプログラムの稼働を担っています。
- ウェブ3層においては、クライアントとなるPCやスマートフォンからの要求は、ウェブサーバ→アプリケーションサーバ→データベースサーバの順で伝達されます。
- 多くのウェブサービスの背後にはデータベースサーバが存在し、クライアントからの要求に応じて、データベース内の「検索」や「読込・書込」をすることで、ウェブサービスが機能しています。
  - ウェブ掲示板やSNS上の投稿が表示され、新たな投稿を記入できるのも、ネットショッピングで過去の購入履歴が表示されるのも、ウェブサーバの背後でデータベースサーバが機能しているためです。

ウェブ3層の仕組み（ネットショッピングサイトにおける事例）



ネットショッピングサイトにおける事例	商品や購入結果の表示	購入総額や送料の計算	請求額データの読み書き

□ 一般のウェブサービスの利用者も、気がつかないうちにデータベースサーバの恩恵を受けています。

サーバの安定した継続稼働とサーバ構築・運営費用の関係

◆ウェブサービスの安定した継続稼働とサーバ運営費用の節約は、両立が困難な問題です。

- ウェブサービスにアクセスが増えて、データベースサーバに高負荷がかかるケースを想定して、それに耐えられる設計にする必要があります。
  - 次第にサービス利用者が増えて行くケースもあれば、メディア等でウェブサービスが紹介されて一時的にサービス利用者が増えるケースもあります。
- 常時稼働しているウェブサービスでは、サーバの処理能力や記録容量が足りない場合でも、一時的にサーバが関わるサービスを停止して、高性能の機器へ交換したり、機器を追加したりすることが困難です。
  - 人気のあるウェブサイト、ウェブサービスは年中無休で24時間アクセスできることが当たり前になっています。
- 一方で、サービス利用者が少ない時に、十二分のサーバ機器で対応することは余分な費用をかけてしまいます。
  - サーバの安定した継続稼働の実現には、サーバの運営費用が増加するジレンマがあります。

利用者数が少ないうちは、1台のPCをサーバコンピュータとして利用するだけでも、負荷に耐えられる。

機器の交換、追加

利用者数が多くなると、高負荷に耐えられるように複数のコンピュータで処理するなど、1台あたりの負荷を軽減する必要がある。

サーバの安定した継続稼働と構築・運営費用の節約のジレンマ

- 一時的にウェブサービスを停止して機器を交換・追加することは避けたい。
- 負荷が小さい時に十二分なサーバ設備を構築、運営するには追加的な費用がかかる。

このジレンマを解決する一つの方法として、**仮想化技術**の活用が挙げられます。

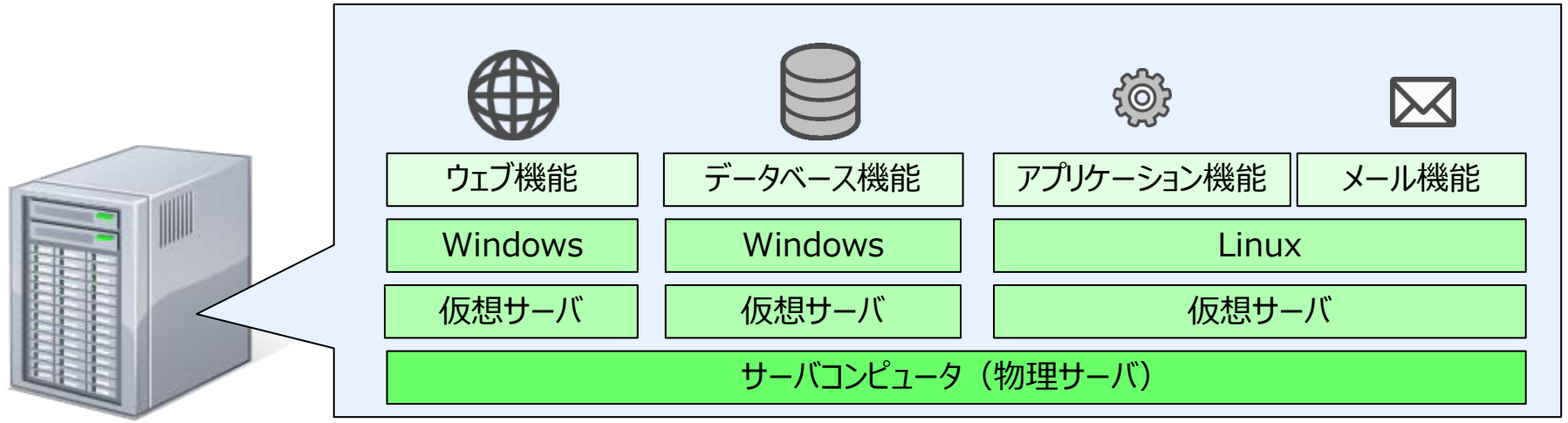
□ クラウドの基本技術でもある仮想化は、サーバの安定した継続稼働とサーバの構築・運営費用の節約に貢献します。

仮想化技術

◆仮想化技術は「物理的なコンピュータの枠にとらわれない」技術です。

- 手元のPCやスマートフォンを操作する場合は、各コンピュータで1つのOSが稼働していることが一般的です。
  - OS（Operating System）は、Windows、Linuxといったコンピュータを動かすための基本ソフトウェアを指しています。
- **仮想化技術**を使うと、1台の物理的なコンピュータの中で複数のOS（例えば、WindowsとLinux）を同時に動かすことが可能です。
  - 代表的な仮想化技術の使い方として、1台の物理的なコンピュータの中で複数の独立した環境を構築することが挙げられます。
  - 仮想化技術は、物理的に複数のコンピュータや機器を1台に見せる技術も含みますが、ここでは仮想サーバの代表的な用途を紹介しています。
- 仮想化技術を使うと、1台のサーバコンピュータ（物理サーバ）の中で独立した環境を分けて、複数の機能を持つ**仮想サーバ**を構築することができます。
  - 物理的に1台、2台と数えられるサーバコンピュータを物理サーバ、仮想化技術で構築した独立したサーバ環境を仮想サーバといいます。
  - 1台の物理サーバ内に複数のOSを稼働させることができ、物理サーバ内の1つのOSを再起動する場合でも、他のOSへの影響はありません。

1台のサーバコンピュータ（物理サーバ）の中の仮想化の事例イメージ

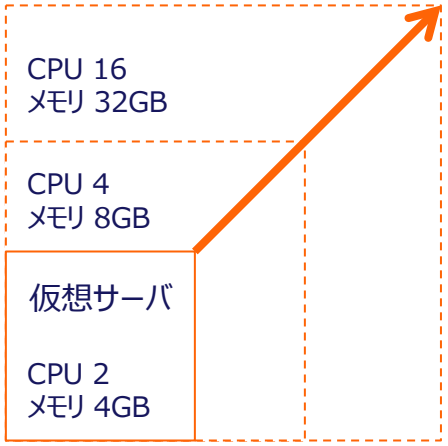


仮想サーバにおけるスケーラビリティ（拡張性）

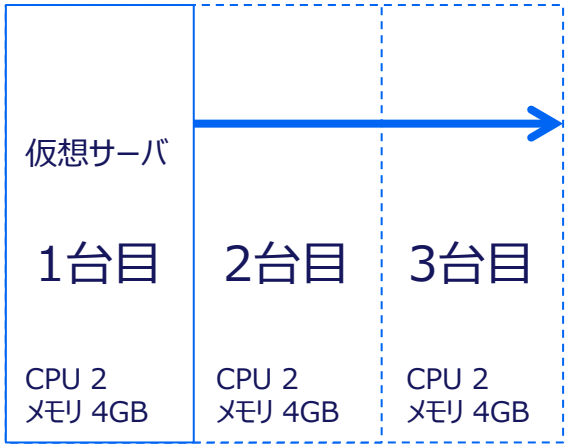
◆仮想サーバには、利用するコンピューターリソースを増減できるスケーラビリティがあります。

- 仮想サーバにおいては、サーバに掛かる負荷に応じて、1つの仮想サーバの性能を高める**スケールアップ**、対応する仮想サーバの台数を増やす**スケールアウト**が可能です。
- スケールアップでは、その仮想サーバが利用できるCPUやメモリの性能を高めることができます。
  - スケールアウトでは物理サーバの内外を問わずに、対応する仮想サーバの台数を増やすことができます。
- 「スケールアップ/スケールダウン」または「スケールアウト/スケールイン」ができることを総じて「**スケーラビリティ（拡張性）**がある」といいます。
- 仮想サーバの中には、サーバにかかる負荷に応じて自動でスケールアップ/スケールダウンを行う**オートスケール**ができるものもあります。

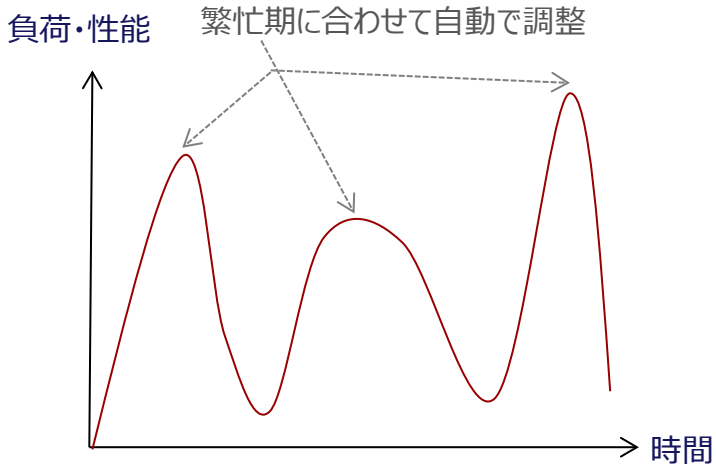
スケールアップ



スケールアウト



オートスケール



□ 仮想サーバのスケーラビリティによって、「サーバの安定した継続稼働とサーバ構築・運営費用の節約」を両立できます。

クラウド コンピューティング

◆クラウドコンピューティングは、仮想化技術を利用したネットワーク関連サービスです。

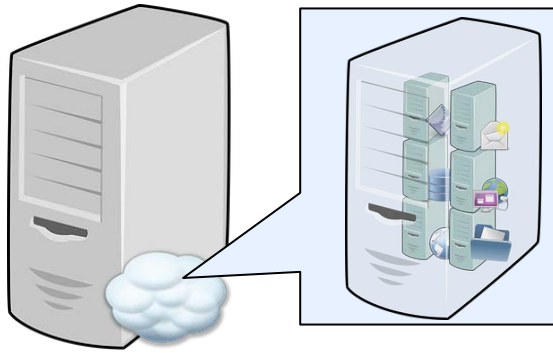
- 仮想化技術は、近年着目されている**クラウドコンピューティング**にも利用されています。
  - 「クラウドコンピューティング」および、それを利用した「クラウドサービス」は、雲を意味する「クラウド（cloud）」と略されることもあります。
- ICT関連用語を平易な文章で説明している総務省が運営する「国民のための情報セキュリティサイト」の用語説明では「クラウドコンピューティング」を次のように記述しています。
  - クラウドコンピューティングに関する用語説明の引用内の青い太文字は、本教材によって記したものです。

「国民のための情報セキュリティサイト」における「クラウドコンピューティング」の用語説明

「インターネット上のネットワーク、サーバ、ストレージ、アプリケーション、サービスなどを共有化して、サービス提供事業者が、利用者に容易に利用可能とするモデルのことです。クラウドコンピューティングには**主に仮想化技術**が利用されています。」

【出所】国民のための情報セキュリティサイト［総務省］
http://www.soumu.go.jp/main_sosiki/joho_tsusin/security/glossary/02.html#ku

クラウドコンピューティングにおける仮想化の事例



仮想技術を用いて、物理的には1台のサーバコンピュータ(物理サーバ)の中に様々な役割を持つ複数の仮想サーバが入っているケースもあります。

- **クラウド コンピューティング**にとって、**仮想化（技術）**は重要なキーワードとなっています。

クラウド コンピューティングの出自

◆「クラウド コンピューティング」は、GoogleのCEOの発言から広まりました。

- インターネットにおける**「クラウド（cloud：雲）」**という言葉は、2006年のGoogleのCEO（最高経営責任者）のEric Schmidt（エリック・シュミット）氏の発言から広がっていきました。

2006年8月の［Search Engine Strategies Conference］におけるEric Schmidt氏の発言

We call it **cloud computing** – they should be in a **"cloud"** somewhere. And that if you have the right kind of browser or the right kind of access, it doesn't matter whether you have a PC or a Mac or a mobile phone or a BlackBerry or what have you – or new devices still to be developed – you can get access to the **cloud.**

【簡潔な日本語訳】「正しい利用方法であれば、OSや機器の種類は問わずに、**その"クラウド（雲）"**にアクセスできる」といったものを、我々は**クラウドコンピューティング**と呼んでいる。

【出所】Search Engine Strategies Conference［Google］ https://www.google.com/press/podium/ses2006.html

- 「クラウドコンピューティング」という言葉は脚光を浴びましたが、「専門家の間でも定義が異なる言葉」「定義や意味が不明確なまま使われている言葉」として、問題視されていました。
  - 2008年12月には米国電気電子学会［IEEE（アイトリプルイー）］の発表においても、専門家の間でも様々な定義があることが示されました。

【出所】Cloud Computing and Grid Computing 360-Degree Compared［IEEE| Ian Foster *et al.* ］
https://ieeexplore.ieee.org/document/4738445/

- 問題解決のため、**米国国立標準技術研究所［National Institute of Standards and Technology（NIST：ニスト）］**は、**「クラウドに必須の5つの特徴」**という資料の草稿を2009年に発表しました。
  - NISTによる草稿は、専門家からの意見を受け付けた後2011年に最終化され、（独）情報処理推進機構によって日本語化がされています。

【出所】NIST によるクラウドコンピューティングの定義［情報処理推進機構］ https://www.ipa.go.jp/files/000025366.pdf

□「クラウドコンピューティング」の意味が不明確な状況は、世界においても、日本においても現在でも継続しています。

NISTが示すクラウドの基本的な特徴

◆NISTは、クラウド（コンピューティング）が満たすべき5つの基本的な特徴を示しています。

- 米国国立標準技術研究所（NIST）が示した**クラウドコンピューティングの5つの特徴**は、表のようにまとめられます。

NISTによるクラウドコンピューティングの5つの特徴（本教材における補足説明付）

番号	NISTが示した基本的な特徴	NISTによる特徴の説明（抜粋）	本教材による補足説明
[1]	**オンデマンド・セルフサービス** (On-demand self-service)	利用者は、クラウド事業者とコミュニケーションの必要なく、サービスを設定できる。	サービスを開始、設定する際に対面や電話での事業者とのやりとりが不要です。
[2]	**幅広いネットワークアクセス** (Broad network access)	サービスはネットワークを通じて利用可能で、標準的な仕組みで接続できる。	インターネット等を通じて、PCやスマートフォンのブラウザ等の一般的な機器で利用できます。
[3]	**リソースの共用** (Resource pooling)	システムリソースは集積され、複数の利用者に提供され得る。	同一の物理サーバを利用しているケースでも、複数の利用者が同時に利用できます。
[4]	**スピーディな拡張性** (Rapid elasticity)	システムリソースは、需要に応じて即座に拡大・縮小できる。	仮想化技術等を活用し、ごく短時間でのスケーラビリティ（拡張性）があります。
[5]	**サービスが計測可能であること** (Measured Service)	システムリソースの利用状況はモニタされ、利用者にもクラウド事業者にも明示できる。	使用したシステムリソース（計算量や記録量）に応じて、課金することが可能になります。

【出所】The NIST Definition of Cloud Computing［NIST］に基づき作成 https://csrc.nist.gov/publications/detail/sp/800-145/final

- NISTが2011年に最終化した本資料では、5つのクラウドの特徴を示したことに加え、講座2-2で示す「Service Model（サービスモデル）」および「Deployment Model（実装モデル）」の分類も表しています。
- NISTは2017年に本資料の5つの特徴に対応するチェックシートを公表し、パブリックコメントを募りました。

【出所】NIST Cloud Computing Program［NIST］ https://www.nist.gov/programs-projects/nist-cloud-computing-program-nccp

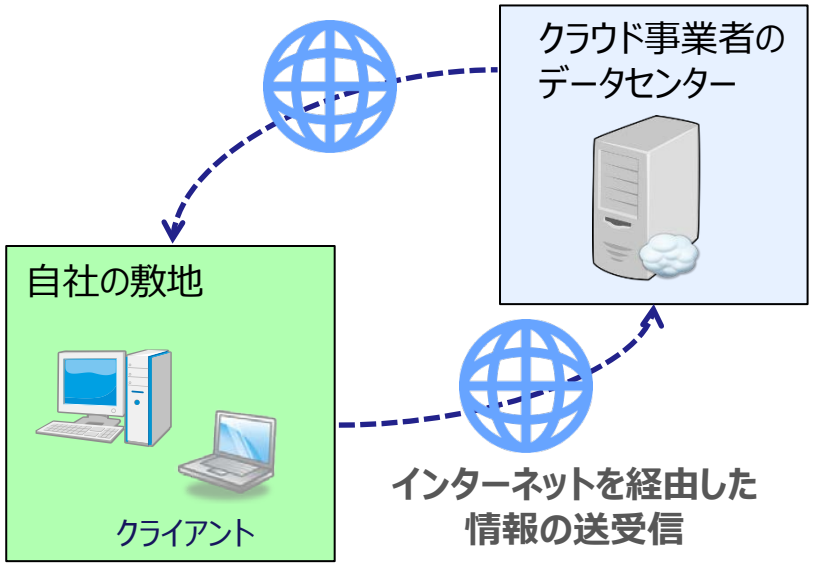
- 表における**［4］の「スピーディーな拡張性」**が、クラウドコンピューティングの際立った特徴となっています。
  - 表における［1］［2］［3］［5］の特徴に関しては、クラウドコンピューティングが脚光を浴びる以前のホスティングサービス（レンタルサーバ）においても一般的に備わっていました。

クラウドとオンプレミス

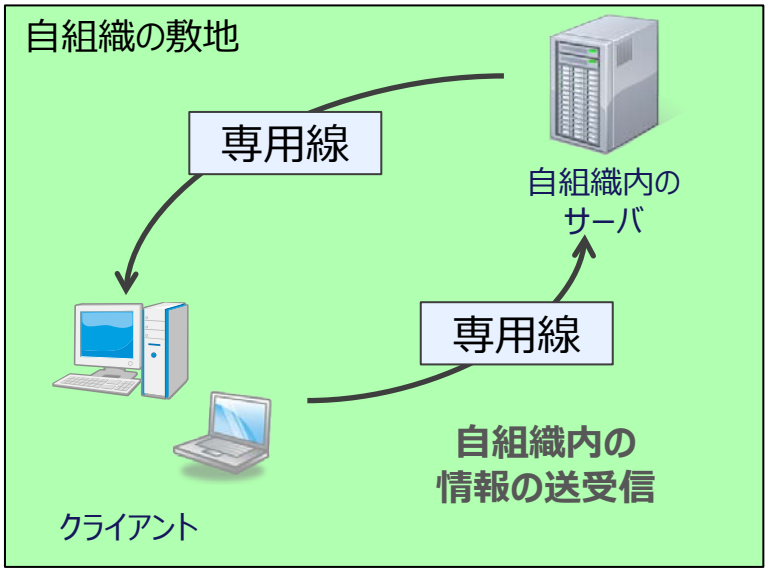
◆クラウドと対照的に紹介されるサーバの運用方法として、オンプレミスが挙げられます。

- クラウドを利用する場合、**インターネットを経由してサービスやデータにアクセスすることが一般的**です。
  - 前スライドの「［2］幅広いネットワークアクセス」は、インターネット接続を必須としていませんが、大半のクラウドはインターネットに接続しています。また、パブリッククラウドと呼ばれる形態ではインターネット接続が必須です。
  - プライベートクラウドと呼ばれるクラウドの形態では、自組織の敷地内に物理サーバを設置するケース（⇒オンプレミス）もあり得ます。パブリッククラウド、プライベートクラウドは講座2-2で詳しく説明します。
- **オンプレミス（on-premise）**と呼ばれる**自組織の敷地内でサーバを運用する形態**では、情報の送受信でのインターネットの利用は不要で、データは自組織の敷地内のサーバに保存されます。
  - 必ずしも建物としての自組織内とは限らず、データセンターの一部を間借り（ハウジング）して、サーバの運営を行うケースもオンプレミスと呼ばれます。

クラウドのイメージ



オンプレミスのイメージ

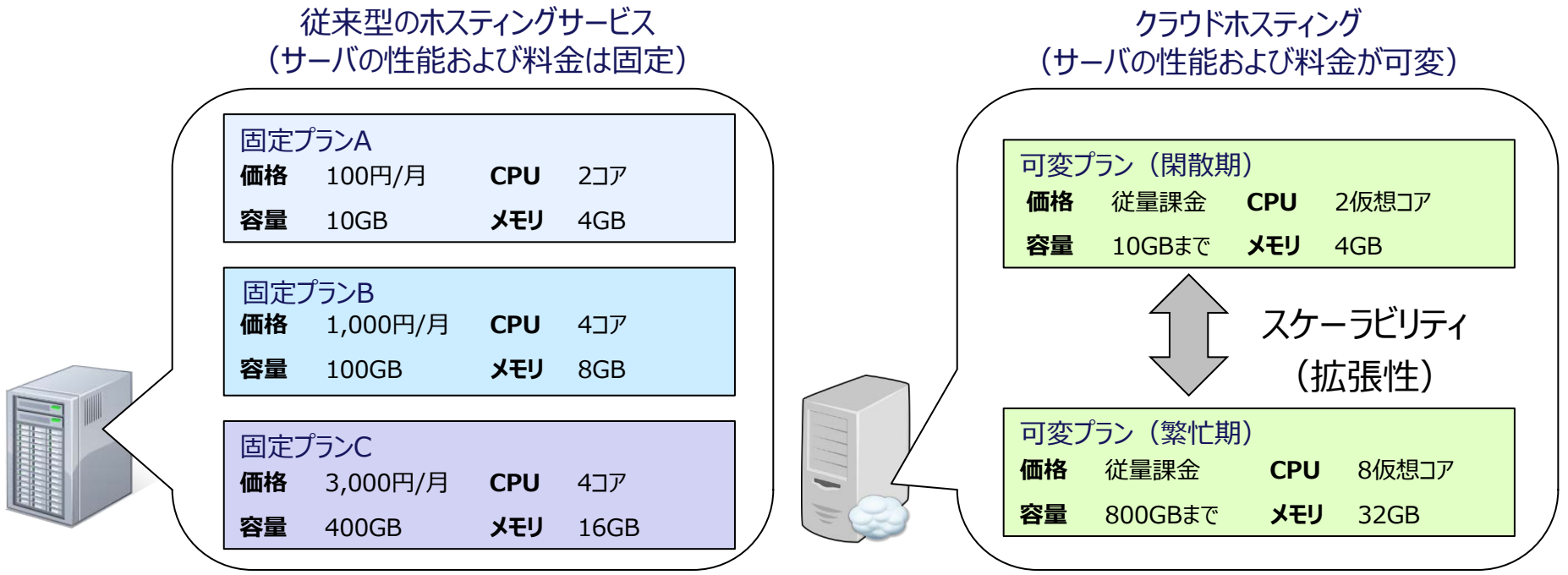


- 自組織の敷地内に物理サーバを設置する**オンプレミス型のクラウドもあり得る**ため、**「クラウド」と「オンプレミス」は対義語ではありません**が、サーバの運用形態として「クラウド」と「オンプレミス」は対照的に紹介されるケースがあります。

クラウドとホスティング

◆ホスティングサービスの中にも仮想化技術を利用した「クラウドホスティング」が登場しました。

- **ホスティングサービス**と呼ばれるデータセンター内の**サーバの利用権を貸し出す仕組み**があります。
  - 「ホスティングサービス」は「レンタルサーバ」とも呼ばれ、クラウドが脚光を浴びる以前から普及していたサービスであり、ウェブサイトの設置等に利用されます。
  - クラウドサービスの中には、仮想サーバの利用権を貸し出すケースもあり、ホスティングサービス同様にウェブサイトの設置にも利用できます。
- 近年新たに登場した**クラウドホスティング**は、クラウドのスケーラビリティ（拡張性）を活用して、利用しているサーバの性能や保存可能な記憶量が可変となります。
  - 一般的なホスティングサービスでは、契約時にサーバの性能および月間または年間の利用料金が固定されている一方で、クラウドホスティングは利用料金が利用している性能に応じて従量制になっているケースが一般的です。

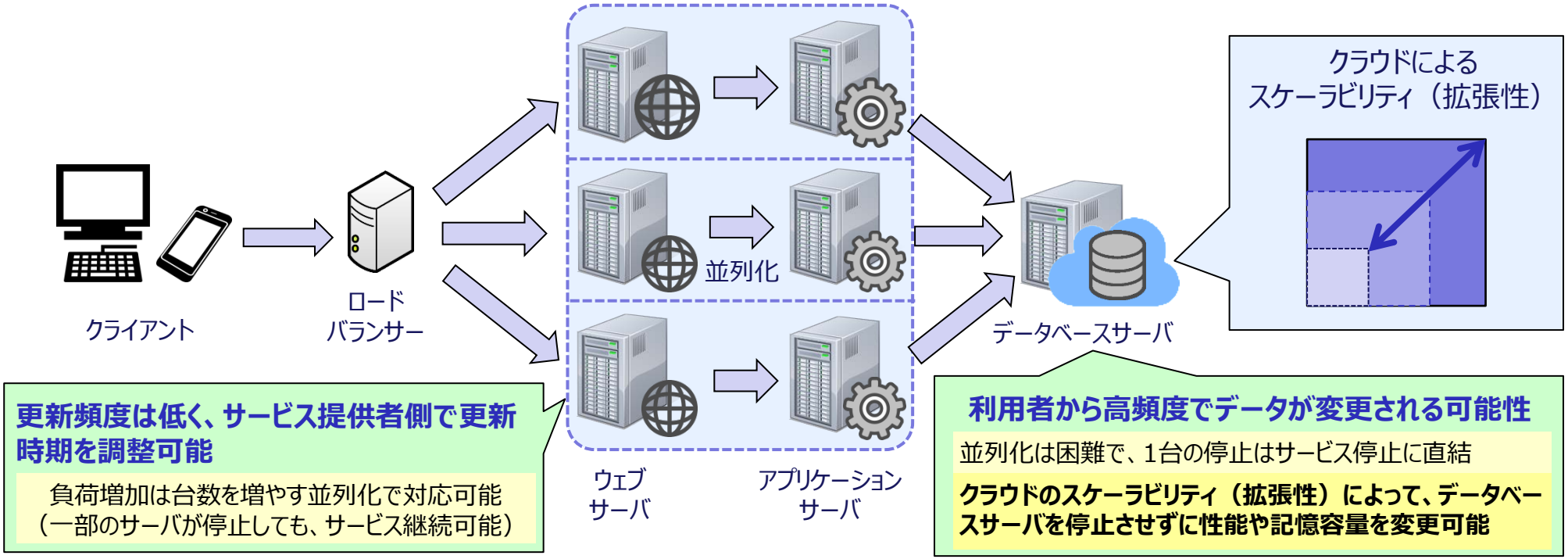


データベースサーバのクラウド利用

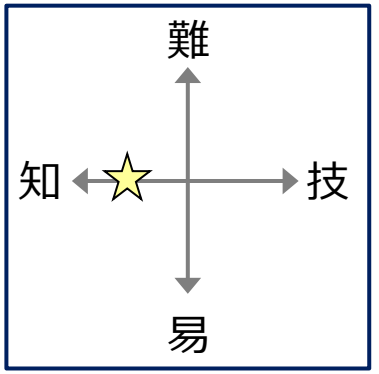
◆クラウドのスケーラビリティ（拡張性）は、データベースサーバへの恩恵が大きくなっています。

- ウェブ3層において、ウェブサーバとアプリケーションサーバは、ウェブサービスの提供者が格納されているデータの更新時期を調整できるため、同一のサーバの台数を増やす**並列化**による負荷分散が容易です。
  - 並列化後のウェブサービスへのアクセスは、ロードバランサーと呼ばれる機器によって、各アクセスに対応するサーバが振り分けられます。
- 一方で、データベースサーバは利用者による高頻度のデータ更新があるため、並列化が困難です。
- クラウドのスケーラビリティ（拡張性）は、並列化が困難なデータベースサーバに特に大きな恩恵をもたらすものとなっています。

ウェブ3層における並列化とデータベースサーバのクラウド利用のイメージ



総務省　ICTスキル総合習得教材



［コース2］データ蓄積

2-2：クラウドのサービスモデル･実装モデル

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_2_2.pdf

	1	2	3	4	5
［コース1］データ収集					
［コース2］データ蓄積		●			
［コース3］データ分析					
［コース4］データ利活用					

本講座の学習内容［2-2：クラウドのサービスモデル・実装モデル］

【講座概要】

- NISTが示したクラウドの2種類の分類の「サービスモデル」および「実装モデル」を示します。
- クラウドの「サービスモデル」を事例を挙げて説明します。
- クラウドの「実装モデル」の分類と事例を説明します。
- 日本企業におけるサービスモデル別・実装モデル別のクラウドの利用状況を把握説明します。

【講座構成】

座学

- [1] クラウドに関する2種類の分類
- [2] 役割分担に基づくサービスモデル
- [3] クラウドのサービスモデルの事例
- [4] 利用機会の開かれ方に基づく実装モデル
- [5] 企業のクラウドの利用状況と市場規模

【学習のゴール】

- ✓ クラウドの「サービスモデル」による分類と各事例を紹介できる。
- ✓ クラウドの「実装モデル」による分類と事例を理解する。
- ✓ 日本企業におけるサービスモデル別・実装モデル別のクラウドの利用状況を把握する。

クラウドの定義に見られる揺れ

◆クラウドには様々な形態があるため、その定義が不明瞭になっています。

- 講座2-1では、「クラウド」という言葉が脚光を浴びたことが一因となり、定義が不明瞭になっていることを示しました。
- クラウドの定義が不明瞭となっている、より大きな理由として**クラウドには様々な形態があり、簡潔な定義が困難であること**が挙げられます。
- NIST(米国国立標準技術研究所）の2009年の公表資料においても、クラウドを明確に定義せず、「必須の5つの特徴」を挙げるに留まりました。
  - NISTの公表資料においても、「クラウドとは、○○である」と簡潔・明瞭な定義はなく、「5種類の特徴を併せ持つもの」と遠回しな表現でした。

NISTによるクラウド（コンピューティング）の5つの特徴［講座2-1の一部を再掲］

番号	基本的な特徴	特徴の説明（抜粋）
[1]	**オンデマンド・セルフサービス** (On-demand self-service)	利用者は、クラウド事業者とコミュニケーションの必要なく、サービスを設定できる。
[2]	**幅広いネットワークアクセス** (Broad network access)	サービスはネットワークを通じて利用可能で、標準的な仕組みで接続できる。
[3]	**リソースの共用** (Resource pooling)	システムリソースは集積され、複数の利用者に提供され得る。
[4]	**スピーディな拡張性** (Rapid elasticity)	システムリソースは、需要に応じて即座に拡大・縮小できる。
[5]	**サービスが計測可能であること** (Measured Service)	システムリソースの利用状況はモニタされ、利用者にもクラウド事業者にも明示できる。

【出所】The NIST Definition of Cloud Computing［NIST］https://csrc.nist.gov/publications/detail/sp/800-145/final

- この講座では、同じNISTの公表資料に示されたクラウドに関する2種類の分類を説明します。

クラウドに関する2種の分類

◆NISTが示したクラウドに関する2種類のモデルは、サービスモデルと実装モデルです。

- NIST(米国国立標準技術研究所）は、2009年の公表資料でクラウドを分類する観点として、**サービスモデル（Service model）**と**実装モデル（Deployment model）**を示しました。
  - この資料はNISTがクラウドの5つの特徴を示した資料と同一の資料であり、NISTではクラウドの5つの特徴を示すとともにクラウドの分類を示しました。
- NISTの当該資料は2011年に最終化され、（独）情報処理推進機構によって、日本語に翻訳されています。

【出所】NIST によるクラウドコンピューティングの定義-米国国立標準技術研究所による推奨-［情報処理推進機構］
https://www.ipa.go.jp/files/000025366.pdf

- サービスモデル（Service model）は、クラウドサービスの構築・カスタマイズに関する**役割分担による分類**です。
  - サービスモデルは「クラウド事業者に任せる範囲」と「利用者が管理・構築する範囲」によって区分されます。
- 実装モデル（Deployment model）は、クラウドサービスの**利用機会の開かれ方による分類**です。
  - 実装モデルは利用機会の公開の程度・公開状態の組み合わせによる区分によって区分されます。

NISTが公開した2種類のクラウドの分類

モデル名	分類の基準	区分
サービスモデル （service model）	クラウドサービスの構築・カスタマイズに関する役割分担による分類	「クラウド事業者に任せる範囲」と「利用者が管理・構築する範囲」による区分
実装モデル （Deployment model）	クラウドサービスの利用機会の開かれ方による分類	利用機会の公開の程度・公開状態の組み合わせによる区分

□ この講座では、まずサービスモデルの分類を説明し、続いて実装モデルの分類を説明します。

クラウドサービスに関する3種の役割分担

◆クラウド事業者と利用者の間の役割分担として3つのケースが考えられます。

単に「クラウドサービス」と言った場合、人や状況によって

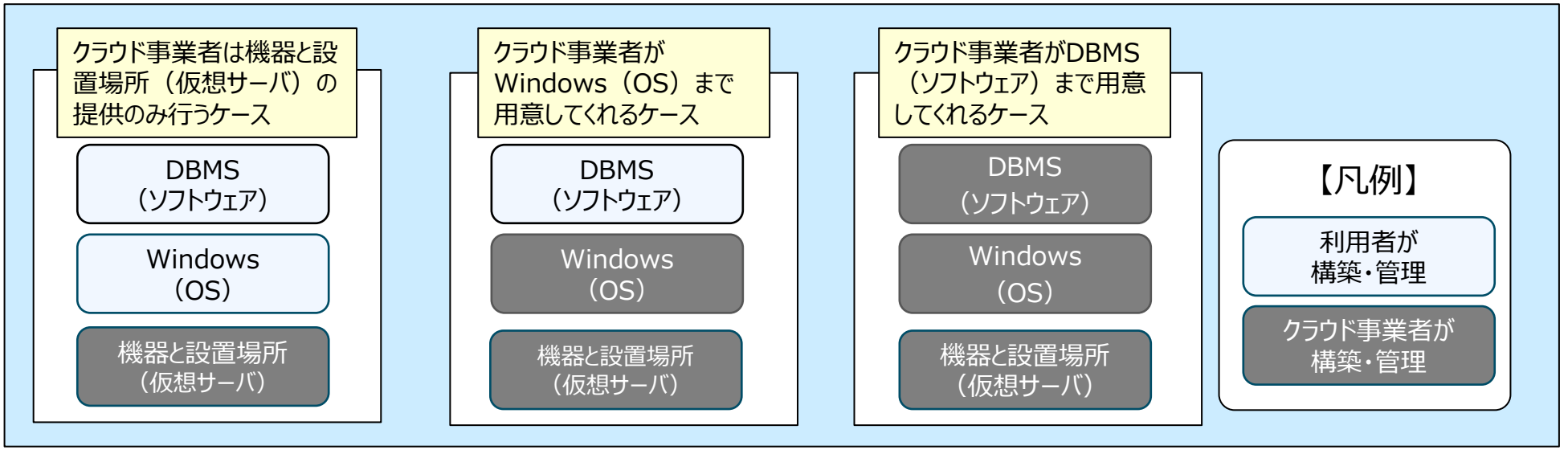
(1) IT企業等が利用する仮想サーバのようなクラウドサービス
(2) 個人が利用するGmailやGoogleドライブのようなクラウドサービス

といった形で、連想するサービスが異なります。

⇨ このようにクラウドサービスとして連想するものが異なるのは、「サービスモデル」の相異にも起因しています。

また(1)の仮想サーバの中で、WindowsをOSとするクラウドサーバにDBMSに設置する場合でも、クラウド事業者と利用者の役割分担として、以下の3種類が考えられます。

クラウドサーバの3種の役割分担（WindowsにDBMSを設置して、データベースサーバとして利用する場合）



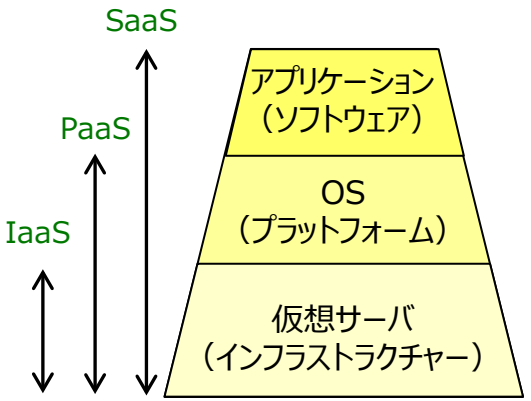
□ これらの3種のクラウド事業者と利用者の役割分担が、クラウドのサービスモデルの分類の考え方につながります。

NISTが示した3種のサービスモデル

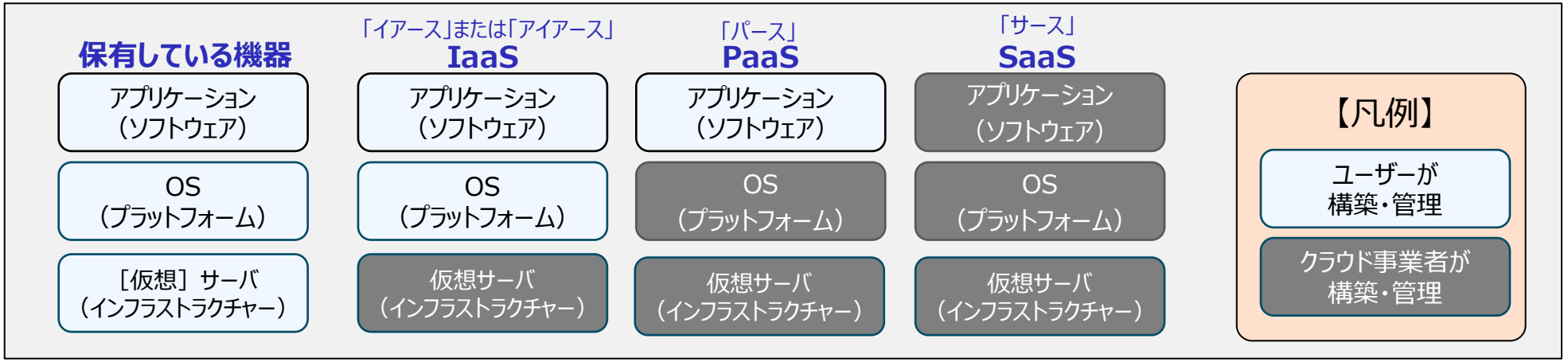
◆サービスモデルは「クラウドサービスの構築に関する役割分担」によって分類されます。

- **サービスモデル**は**クラウドサービスの構築・カスタマイズに関する役割分担**によって、下記の3種に分類されます。
  - ◆ **IaaS**（イアースまたはアイアース） ［Infrastructure as a Service：サービスとして提供されるインフラストラクチャー］
  - ◆ **PaaS**（パース） ［Platform as a Service：サービスとして提供されるプラットフォーム］
  - ◆ **SaaS**（サース） ［Software as a Service：サービスとして提供されるソフトウェア］
    - 「インフラストラクチャー」は「基盤・下支えするもの」を指し、クラウドサービスにおいては仮想サーバを指しています。
    - プラットフォームとは、個別の作業をするアプリケーション（ソフトウェア）を導入する土台となる基本ソフトウェアであり、OSを含みます。
- コンピュータの階層を3層で考えた場合、1段目のみクラウド事業者に任せるのがIaaS、2段目まで任せるのがPaaS、3段目を含めて全て任せるのがSaaSです。

コンピュータの階層と任せる範囲



保有機器および各クラウドのサービスモデルの「構築・管理範囲」



□ クラウドのサービスモデルの区分によって、どこまでクラウド事業者に準備してもらえるかの範囲が異なります。

クラウドのサービスモデルと主なサービス利用者

◆クラウドのサービスモデルの分類は、主な利用者の分類にも通じています。

- 2017年のNISTの公開資料において、サービスモデルによる分類は、クラウドサービスの主たる利用者にも対応していると論じています。

【出所】DRAFT - Evaluation of Cloud Computing Services Based on NIST 800-145［NIST］
https://www.nist.gov/sites/default/files/documents/2017/05/31/evaluation_of_cloud_computing_services_based_on_nist_800-145_20170427clean.pdf

クラウドの各サービスモデルと主な利用者

略称	カタカナでの発音	正式名	日本語訳	主な利用者
IaaS	「イアース」 または「アイアース」	Infrastructure as a Service	サービスとして提供される インフラストラクチャー	ICTサービスの運営者
PaaS	「パース」	Platform as a Service	サービスとして提供される プラットフォーム	ソフトウェアの開発者・実装者
SaaS	「サース」	Software as a Service	サービスとして提供される ソフトウェア	一般ユーザー（エンドユーザー）

- エンドユーザーとは、根本的な設定変更やプログラミングを行わずに、構築済のソフトウェアを利用するユーザーを指します。

IaaSの主な利用者：
ICTサービスの運営者

PaaSの主な利用者：
ソフトウェアの開発者・実装者

SaaSの主な利用者：
一般ユーザー（エンドユーザー）

□「クラウドサービス」として連想するものが異なるのは、個人による主な利用サービスの相異にも起因しています。

XaaS（ザース）

◆様々なクラウドサービスの総称として、NISTが定義していないXaaSという概念があります。

- IaaS、PaaS、SaaS以外にも、業界や企業のマーケティング用語としての**XaaS（ザース）**と呼ばれるクラウドサービスがあります、Xの部分には様々なアルファベットが入ります。
  - ICTの技術面では、様々な文字が入る部分に「x」を入れて総称するケースがあります。例えば、講座1-1に示した様々な産業への情報通信技術の適用を指すX-tech（エックステック）、講座1-2に示した様々なタイプの仮想現実を表すxR（エックスアール）が挙げられます。

- NISTは2017年の公表資料において、XaaSは上位概念となる3種のサービスモデルに含まれうる可能性を示唆し、「NISTは現時点において、これらのクラウドサービスについて、定義をする立場をとらない」としています。
  - NISTの公表資料においては、簡単にウェブ検索をした範囲で83種のXaaSが確認されたことを示しています。

NISTの公表資料に示されたXaaSの事例

略称	正式名	本教材による日本語訳
AaaS	Authentication as a Service	サービスとしての認証
BaaS	Backend as a Service	サービスとしてのバックエンド
	Backup as a Service	サービスとしてのバックアップ
CaaS	Communication as a Service	サービスとしてのコミュニケーション
DaaS	Data as a service	サービスとしてのデータ
	Desktop as a Service	サービスとしてのデスクトップ

【出所】DRAFT - Evaluation of Cloud Computing Services Based on NIST 800-145（NIST）
https://www.nist.gov/sites/default/files/documents/2017/05/31/evaluation_of_cloud_computing_services_based_on_nist_800-145_20170427clean.pdf

- この表に示しているように、BaaSやDaaSと略されるサービスの中でも、異なる種類のサービスを指しているケースもあります。

□ 様々なXaaSの登場は、様々なクラウドサービスが生まれていることを示唆しています。

IaaS/PaaS型　クラウドサービス

◆IaaS/PaaSの型の有力なクラウドサービスとして「AWS」「GCP」「Azure」が挙げられます。

- 世界的に有力なクラウドサービスとして、**3大クラウドプラットフォーム**とも呼ばれる、**AWS（エーダブリューエス）**、**Azure（アジュール）**、**GCP（ジーシーピー）**の3種の総合クラウドサービスが挙げられます。
  - AWSは「Amazon Web Service」の略でAmazonが提供、「Azure」はMicrosoft社提供、GCPは「Google Cloud Platform」の略でGoogleが提供しています。本教材においては、それぞれAWS、Azure、GCPと略称で呼称します。
  - これら3種のクラウドサービスは、IaaS/PaaS型のサービスを中心に様々なサービスメニューが用意され、世界規模でのサービス展開を行っています。

3大クラウドプラットフォームと呼ばれる有力な3種のクラウドサービス

Amazon Web Servicesは、米国
その他の諸国における、Amazon.com,
Inc.またはその関連会社の商標です。

Microsoft は、米国 Microsoft Corporation
およびその関連会社の商標です。

© 2017 Google LLC All rights reserved.
Google および Google ロゴは Google LLC の
登録商標です。

- これらのクラウドプラットフォームでは、簡単な操作で仮想サーバを構築できることに加えて、1時間単位などの短時間で利用できるサービスもあります。
  - クラウドの5つの特徴の1つとして「サービスが計測可能であること(Measured Service)」が挙げられ、時間単位や利用したリソースに応じた課金が可能となっています。
  - それぞれのクラウドプラットフォームは、講座2-5にて紹介するGPUクラウドコンピューティングや分散技術を活用したサービス、講座3-5にて紹介する人工知能技術を活用したクラウドサービスも展開しています。
- これらのクラウドプラットフォームでは、IaaS/PaaS型を中心に様々なサービスを提供しています。
  - AWS、Azure、GCPのサービスメニューは数多くのサービスで構成されており、サービスによってIaaSに対応するもの、PaaSに対応するものがあります。

クラウドプラットフォームのサービスカテゴリ

◆クラウドプラットフォームでは、様々なクラウドサービスを提供しています。

- AWS（Amazon Web Service）では、下図のように様々なカテゴリのクラウドサービスを展開しています。
  - 各カテゴリの中では、概ね複数のサービスを示されており、クラウドプラットフォームでは多種多様なサービスを提供しています。

AWS（Amazon Web Service）におけるクラウドサービスのカテゴリの一覧［2018年3月時点］

【出所】クラウドサービスのカ一覧 ［Amazon Web Services, Inc.］
https://aws.amazon.com/jp/products/

- 2018年3月時点において、AWSは19のサービスカテゴリ、101種のクラウドサービスを公開しています。

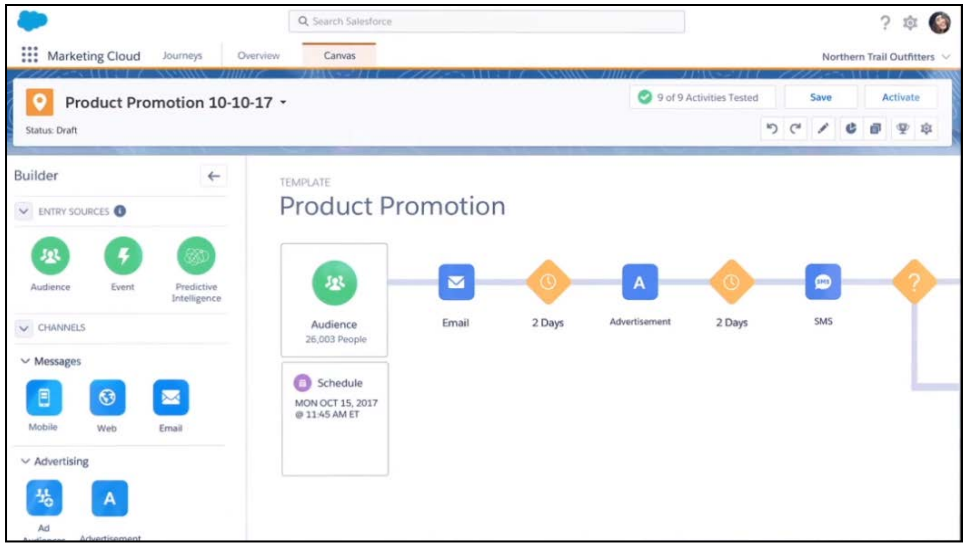
PaaSの事例

◆PaaSでは、ソフトウェア・サービスを開発しやすい環境が、あらかじめ整えられています。

- PaaSでは、業務アプリケーションやモバイル端末向けアプリケーションなどの開発環境があらかじめ整えられています。

- アメリカに本社のあるsalesforce.comが提供するPaaS型クラウドサービスのMarketing Cloudでは、マーケティング（販促活動）のアプリケーションを容易に開発できるようになっています。
  - マーケティングのアプリケーションの一つの事例として下図のようにフローチャートを作る機能があらかじめ備わっています

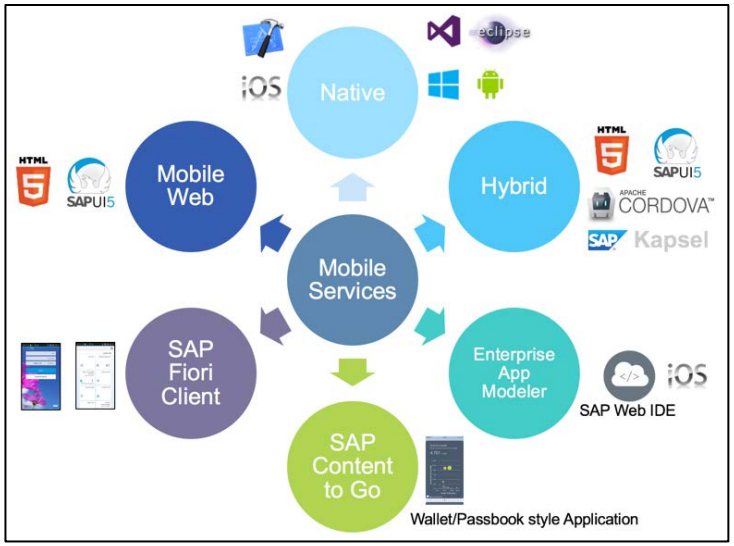
Marketing Cloudにおけるフローチャート作成



【出所】Marketing Cloud［株式会社セールスフォース・ドットコム］
https://www.salesforce.com/jp/products/marketing-cloud/overview/

- ドイツに本社のあるSAPは、PaaS型クラウドサービスのSAP Cloud Platformを提供しています。
- SAP Cloud Platformのサービスの一つとして、モバイルアプリケーションの開発・運用やデバイス管理を容易にするMobile Serviceが挙げられます。

SAP Cloud PlatformのMobile Serviceの構成



【出所】SAP Cloud Platform ［SAPジャパン株式会社］
https://www.sapjp.com/blog/archives/17837

SaaSの事例

◆SaaS型のクラウドサービスとして、電子メールサービスやファイル共有サービスが挙げられます。

- SaaS型クラウドサービスでは、インターネットを活用するサービスとしてソフトウェアを利用できます。
- SaaSの利用には「ウェブブラウザから利用するケース」と「専用のアプリケーションによって利用するケース」があります。
  - ウェブブラウザからSaaSを利用するケースでは、利用者によるソフトウェアの更新（アップデート）作業の必要はありません。
- 代表的なSaaS型クラウドサービスとして、主にウェブブラウザを通じて電子メールサービスを利用するGmailが挙げられます。
- Googleドライブに代表されるファイル保存・共有サービスもSaaS型のクラウドサービスです。

Gmailの利用画面

Googleドライブの利用画面

© 2017 Google LLC, used with permission.　Google および Google ロゴは Google LLC の登録商標であり、同社の許可を得て使用しています。

◘ 以上で、クラウドの「サービスモデル」の説明を終え、続いて「実装モデル」を説明します。

NISTが示した4種の実装モデル

◆実装モデルは、「クラウドサービスの利用機会の開かれ方」によって4種に分類されます。

- クラウドの**実装モデル**は、クラウドサービスの**利用機会の開かれ方**によって、4種に分類されます。
- **プライベートクラウド**は、同一の組織に属する者にのみ利用機会が開かれたクラウドです。
  - プライベートクラウドは、物理サーバを利用組織の敷地内に設置するケースもあれば、クラウド事業者のデータセンターなどに設置するケースもあります。
- **コミニティクラウド**は、特定のコミュニティに属する組織・個人にのみに利用機会が開かれたクラウドです。
- **パブリッククラウド**は、利用機会が一般公開されており、利用規約を承諾し登録すれば、誰でも使えるクラウドです。
  - AWSやGmailは、利用機会が公開されており、登録すれば誰でも使えるパブリッククラウドです。
- **ハイブリッドクラウド**は、複数のクラウドの実装モデルを適材適所に組み合わせる形態です。

クラウドサービスの実装モデルの分類

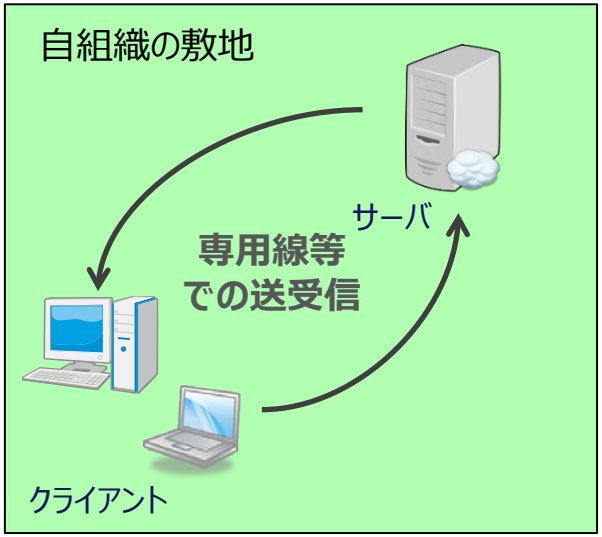
実装モデル	利用権限がある者	サービスの提供者
プライベートクラウド (Private cloud)	同一組織内の部門や個人	その組織自体、またはクラウドの運営を委託された外部組織
コミュニティクラウド (Community cloud)	業界団体や特定のコミュニティ（共同体）に属する組織	コミュニティの構成組織、またはクラウドの運営を委託された外部組織
パブリッククラウド (Public cloud)	登録すれば、誰でも利用可能	クラウドサービスを公に提供しているクラウド事業者
ハイブリッドクラウド (Hybrid cloud)	ハイブリッドクラウドは、上記3種の実装モデルの組み合わせであるため、クラウド内の仮想サーバやディレクトリ（フォルダ）やデータ種によって異なります。	

プライベートクラウド

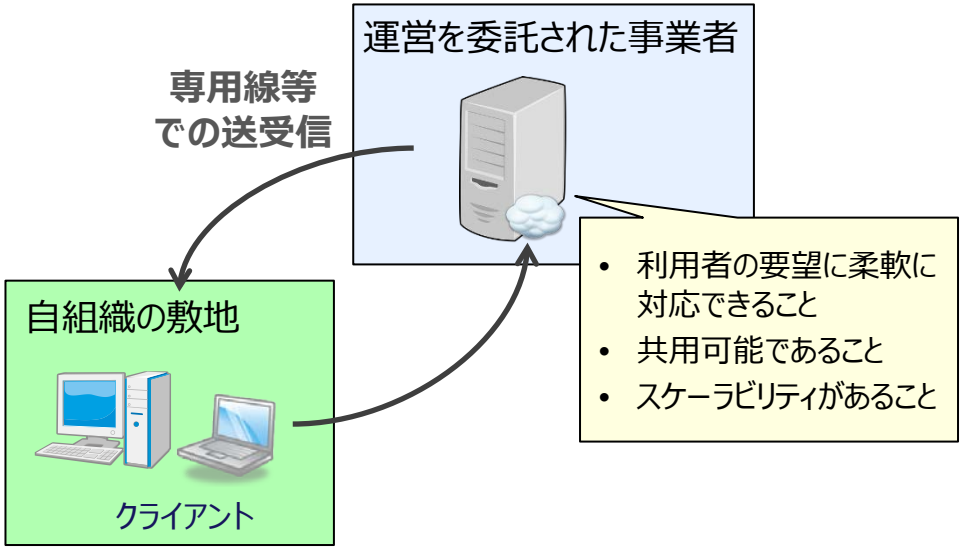
◆プライベートクラウドは、特定の企業や組織が独自に利用するクラウドです。

- プライベートクラウドの利用者は、特定の企業などの同一組織に属する部門や個人です。
  - プライベートクラウドの提供者は、その組織自体または運営を委託された外部組織です。
- プライベートクラウドは、そのサーバを自組織の敷地内に設置する場合もあれば、敷地外に設置する場合もあります。
  - 自組織の敷地内に設置する場合は、講座2-1に示したオンプレミスのプライベートクラウドとなります。
  - 自組織の敷地外に設置するプライベートクラウドは、ホスティング型のプライベートクラウドと呼ばれることがあります。ただし、プライベートクラウドと言えるためには運営を委託された外部組織は、回線やサーバの設計、サービス要望に柔軟に対応できるなど、パブリッククラウドに該当しないことが必要です。
- クラウドであるためには、技術的に共用可能であること、スケーラビリティ（拡張性）があることも条件となります。
  - インターネットでの利用はクラウドの条件として必須ではありませんが、企業内の全社員がアクセスできるなど、幅広いネットワークでの利用ができることもクラウドの条件です。

自組織の敷地内に設置する
（オンプレミス型）プライベートクラウド



自組織の敷地外に設置する
（ホスティング型）プライベートクラウド



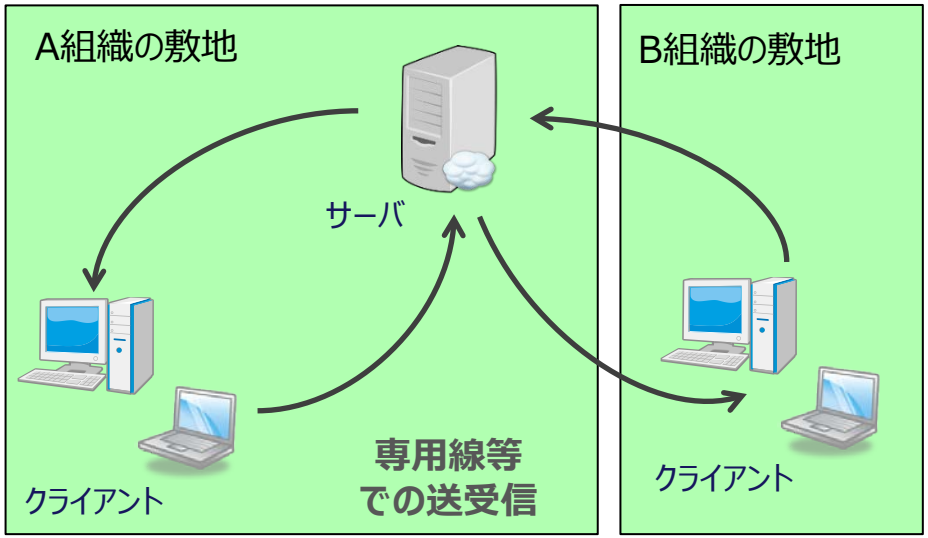
コミュニティクラウド

◆ コミュニティクラウドは、目的や業務が関連する複数組織で共同利用されるクラウドです。

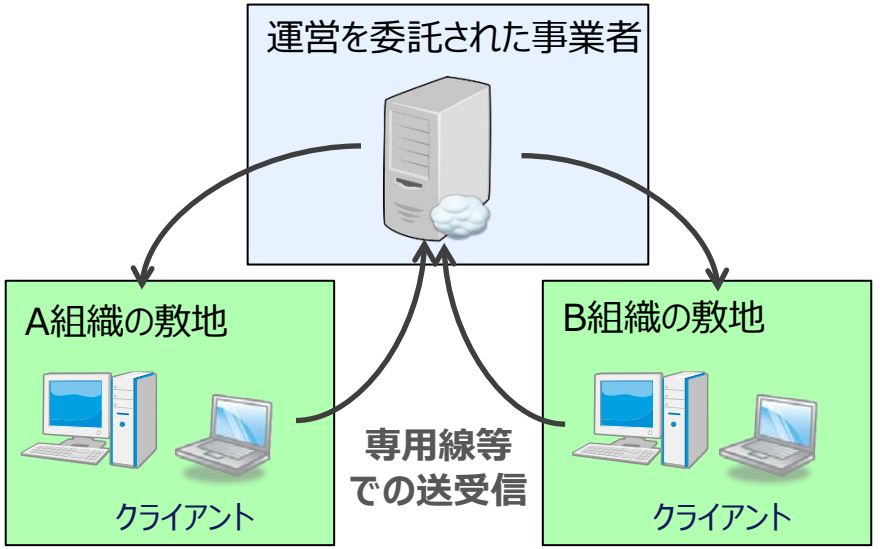
- コミュニティクラウドは、複数の組織、個人で構成される団体など、コミュニティ（共同体）で利用するクラウドです。
  - コミュニティクラウドの提供者は、コミュニティクラウドを構成する組織または運営を委託されたサードパーティです。
- コミュニティクラウドは、特定のコミュニティ内における情報共有や共同作業のために利用されます。
  - 同等の性能のクラウドであれば、プライベートクラウドに比べて一組織あたりの費用は低下します。
- コミュニティクラウドの例として、銀行間の情報共有や金融サービスの連携を目的としたクラウドや各府省が共同で利用する行政クラウドが挙げられます。
  - 総務省行財政管理局が中心となって、2013年より政府情報システムのクラウド化した「政府共通プラットフォーム」を運用しています。

【出所】政府共通プラットフォームにおける取組について［総務省］
http://www.soumu.go.jp/menu_news/s-news/01gyokan05_02000025.html

特定の組織内にサーバを設置するコミュニティクラウド



委託先にサーバーを設置するコミュニティクラウド

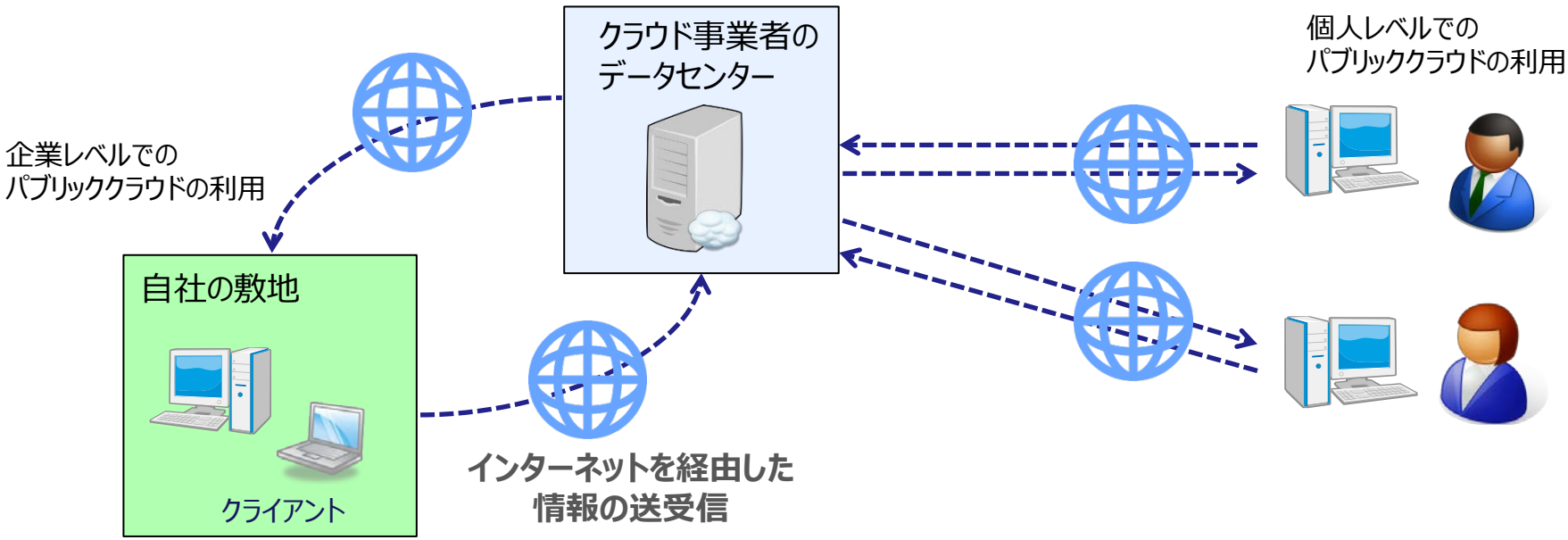


パブリッククラウド

◆パブリッククラウドは、利用機会が公開され、インターネット経由で利用されるクラウドです。

- パブリッククラウドは利用機会が一般公開されており、利用規約を承諾し登録すれば誰にでも利用できるクラウドです。
  - パブリッククラウドは利用できる機会が開かれているということであって、データを公開することを意味しません。
- パブリッククラウドの提供者は、クラウドサービスを実施しているクラウド事業者です。
  - パブリッククラウドの設備はクラウド事業者が構築し、利用者はインターネットを経由した情報の送受信を経てサービスを利用します。
  - パブリッククラウドでは、利用料金やサービス内容が公開されており、クラウド事業者は一般に個々の利用者の要望に応じたカスタマイズを行いません。
- AWS、Azure、GCPといったクラウドプラットフォームのサービスは利用機会が公開されており、利用規約を承諾して登録すれば誰でも利用できるパブリッククラウドに該当します。

パブリッククラウドの利用イメージ



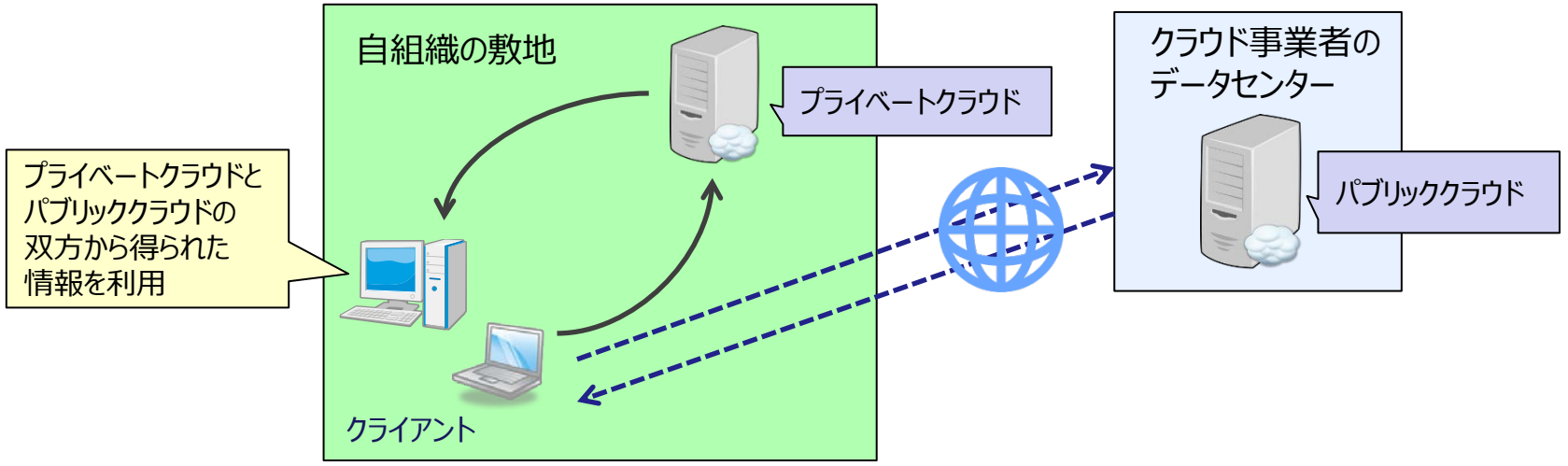
ハイブリッドクラウド

◆ハイブリッドクラウドは、クラウドの実装モデルを組み合わせて利用する形態です。

- ハイブリッドクラウドは、これまでに示した3種の実装モデルをネットワーク内で部分的に組み合わせた形態です。
  - 同じ実装モデルのパブリッククラウド同士であっても、AWSとGmailを組み合わせて利用するなど、複数のクラウドサービスを組み合わせて利用する場合は「マルチクラウド」といいます。
- ハイブリッドクラウドの例として、複雑なカスタマイズが必要な部分はプライベートクラウドで構築し、定型的なサービスで対応できる部分はパブリッククラウド上に構築するケースが考えられます。
- ハイブリッドクラウドを円滑に利用できるよう複数のクラウドをまたぐ認証を一括して行えるサービスもあります。
  - IBMでは、ハイブリッドクラウド対するシングルサインオンサービスとしてXaaSの一種のIdentity-as-a-Service（IDaaS：アイダース）を提供しています。

【出所】IBM、Identity-as-a-Serviceをハイブリッド・クラウド環境に提供［IBM］
https://www-03.ibm.com/press/jp/ja/pressrelease/52649.wss

ハイブリッドクラウドの利用イメージ



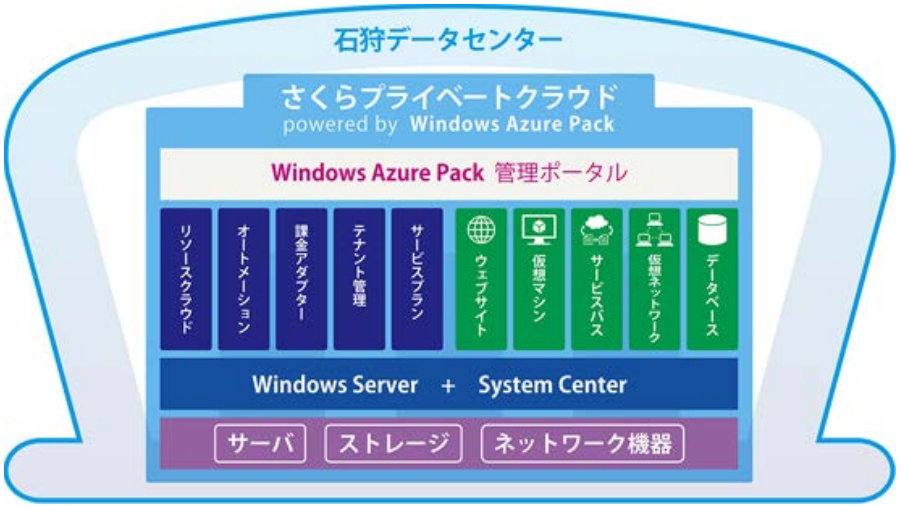
クラウドの実装モデルの事例

◆クラウドの実装モデルはそれぞれの特徴を活かし、様々な形態で提供されています。

- さくらインターネット株式会社では、石狩データセンター内にプライベートクラウドを構築するサービスを提供しています。

- 石狩データセンターのプライベートクラウド管理システムは、MicrosoftのAzureの技術を活用しています。
  - プライベートクラウドとAzureのパブリッククラウドのサービスを連携させるハイブリッドクラウドの構築も可能となっています。

石狩データセンターでのプライベートクラウドのイメージ

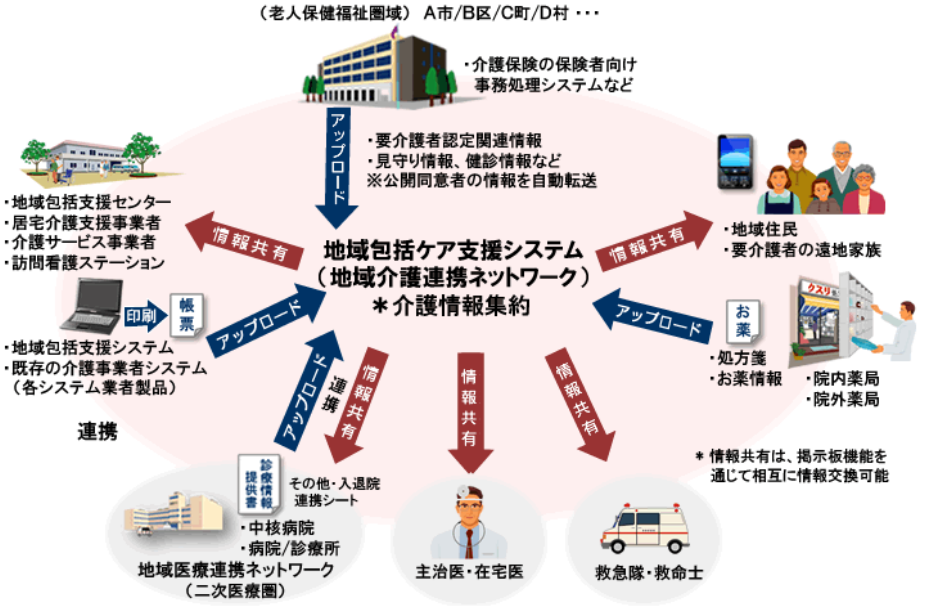


【出所】さくらインターネット、Microsoft Azure基盤を採用した「さくらプライベートクラウド」の提供を開始［さくらインターネット株式会社］
https://www.sakura.ad.jp/press/2015/0903_private-cloud/

- 日立株式会社では、地域包括ケア支援システム（地域介護連携ネットワーク）として、地域の医療機関・介護施設が情報をアップロードできるコミニティクラウドに該当するシステムを提供しています。
  - 地域包括ケアシステムは、厚生労働省が構築を推進する医療・介護に関する地域の包括的な支援・サービス提供体制です。

【出所】地域包括ケアシステム［厚生労働省］
http://www.mhlw.go.jp/stf/seisakunitsuite/bunya/hukushi_kaigo/kaigo_koureisha/chiiki-houkatsu/

地域包括ケア支援自治体クラウドソリューションのイメージ



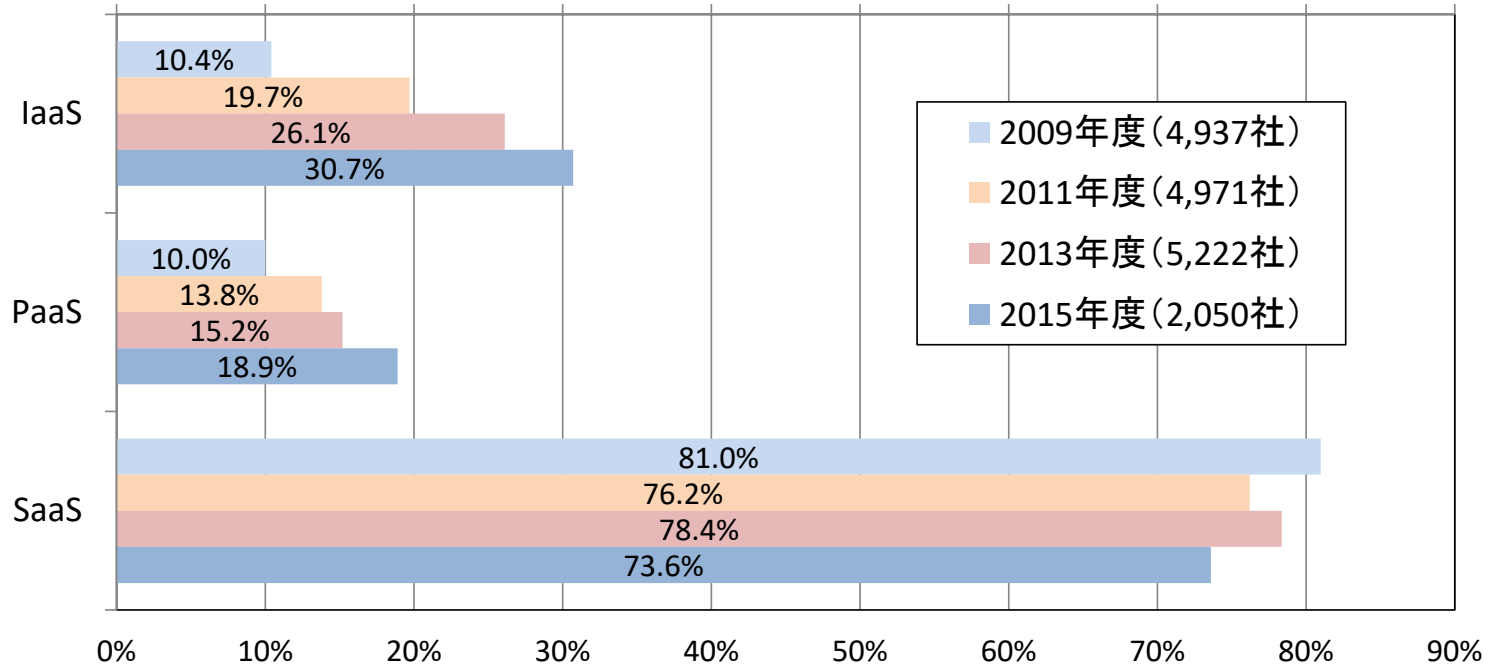
【出所】地域包括ケア支援自治体クラウドソリューション［株式会社日立製作所］
http://www.hitachi.co.jp/Prod/comp/app/chikea/

日本企業におけるサービスモデル別クラウドの利用割合

◆企業によるクラウドの利用割合はSaaSの水準が高く、IaaSやPaaSは増加しています。

- 経済産業省の情報処理実態調査では、日本企業におけるクラウドのサービスモデル別の利用割合を調査しています。
  - 情報処理実態調査において、調査対象としているのは、資本金3000万円以上かつ総従業員50人以上の企業です。
- 2015年度における調査結果では、IaaSが30.7%、PaaSが18.9%、SaaSは73.6%と、SaaSの利用割合の水準が最も高くなっています。
- 推移を見ると、SaaSの利用割合は減少傾向があるのに対して、IaaSおよびPaaSの利用割合は増加しています。

日本企業におけるサービスモデル別クラウドの利用割合の推移



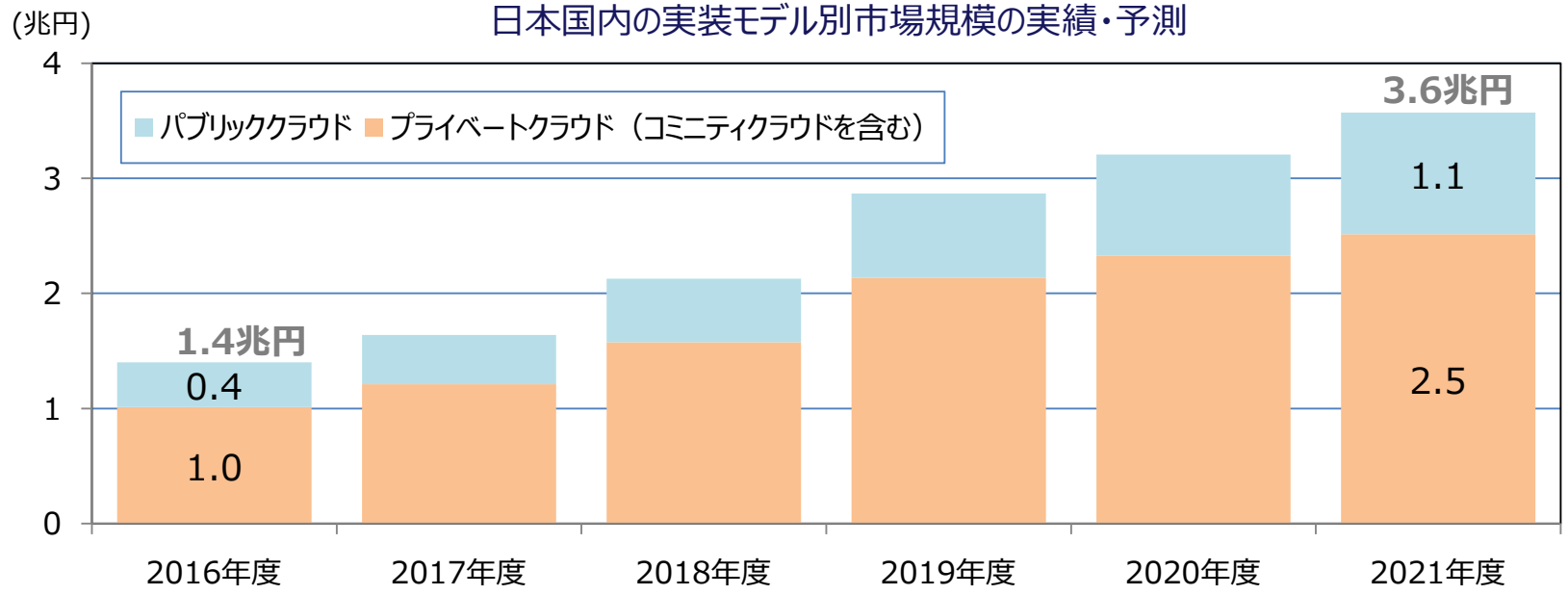
【出所】情報処理実態調査［経済産業省］に基づき作成
http://www.meti.go.jp/statistics/zyo/zyouhou/index.html

実装モデル別クラウドサービスの市場規模の予測

◆日本のクラウドサービス市場は、今後も拡大していくことが見込まれています。

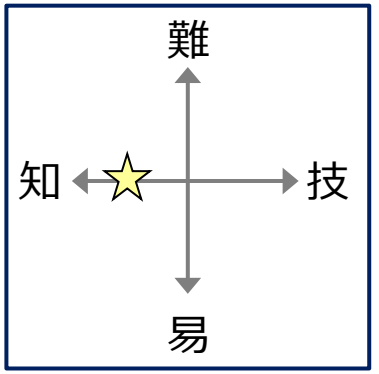
- 株式会社MM総研の推計では、クラウドサービスの国内市場規模が今後拡大していくことが示されています。
  - 株式会社MM総研では、アンケート調査等に基づいてクラウドの実装モデル別の市場規模を予測し、公表しています。
- 2016年度のクラウドサービスの市場規模は約1.4兆円でしたが、2021年度には約3.6兆円が見込まれています。
- 2021年度における市場規模予測の内訳は、パブリッククラウドが約1.1兆円（3割）、プライベートクラウドが約2.5兆円（7割）となっています。
  - MM総研が公表している推計値においては、コミニティクラウドもプライベートクラウドに含まれています。

日本国内の実装モデル別市場規模の実績・予測



【出所】2016年度の国内クラウド市場は4割増の1.4兆円［MM総研］に基づき作成
https://www.m2ri.jp/news/detail.html?id=279

総務省　ICTスキル総合習得教材



［コース2］データ蓄積

2-3：クラウドの特性とセキュリティ

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_2_3.pdf

	1	2	3	4	5
［コース1］データ収集					
［コース2］データ蓄積			●		
［コース3］データ分析					
［コース4］データ利活用					

本講座の学習内容［2-3：クラウドの特性とセキュリティ］

【講座概要】

- 通信利用動向調査（企業調査）を用いて、国内企業におけるクラウドの利用状況を示します。
- 国内企業が「クラウドを利用している理由」「クラウドを利用しない理由」を紹介します。
- 情報セキュリティの重要性および情報セキュリティにおける3つの要素を説明します。
- クラウドサービスに関する情報セキュリティとして「責任共有モデル」を示し、サービスモデル別のクラウド事業者と利用者の責任分担を説明します。

【講座構成】

座学

- [1] 国内企業のクラウドの利用状況
- [2] クラウドを利用している/利用しない理由
- [3] 情報セキュリティの3つの要素
- [4] クラウドと情報セキュリティ

【学習のゴール】

- ✓ 国内企業のクラウド利用状況の推移を理解する。
- ✓ 国内企業が「クラウドを利用している理由」「クラウドを利用しない理由」の概要を把握する。
- ✓ 情報セキュリティの3つの要素を紹介できる。
- ✓ クラウドの「責任共有モデル」を理解する。

クラウドの利用状況と利用に関する理由

◆通信利用動向調査（企業調査）は、国内企業にクラウドの利用状況等を尋ねています。

- 通信利用動向調査は総務省が実施する公的統計であり、情報通信の利用動向を調査しています。
  - 通信利用動向調査には、世帯向け調査と企業向け調査の2種類があります。
- 2016年9月末を調査時点とする平成28年調査（企業調査）は、2,032社の国内企業から回答が得ました。
  - 通信利用動向調査（企業調査）の調査対象となるのは、公務を除く産業に属する常用雇用者規模100人以上の企業です。
- 通信利用動向調査（企業調査）では、調査項目として「クラウドの利用状態」と「クラウド利用に関する選択理由」を調査しています。

平成28年通信利用動向調査（企業調査）におけるクラウド利用に関する質問

問5　貴社におけるクラウドコンピューティング※の利用状況についてお尋ねします。

（1）**クラウドコンピューティング**（以下「クラウド」といいます。）注**を利用**していますか。**該当する番号1つに○印**を付けてください。

1. 全社的に利用している
2. 一部の事業所又は部門で利用している
3. 利用していないが、今後利用する予定がある
4. 利用していないし、今後利用する予定もない
5. クラウドについてよく分からない

注　ネットワーク上に存在するコンピュータ資源を、インターネット等のブロードバンド回線を経由して、利用者が「必要な時に、必要な量だけ」、役務（サービス）として使用できる技術であり、ASP※（Application Service Provider）が提供するSaaS※（Software as a Service）などもこれに含まれます。

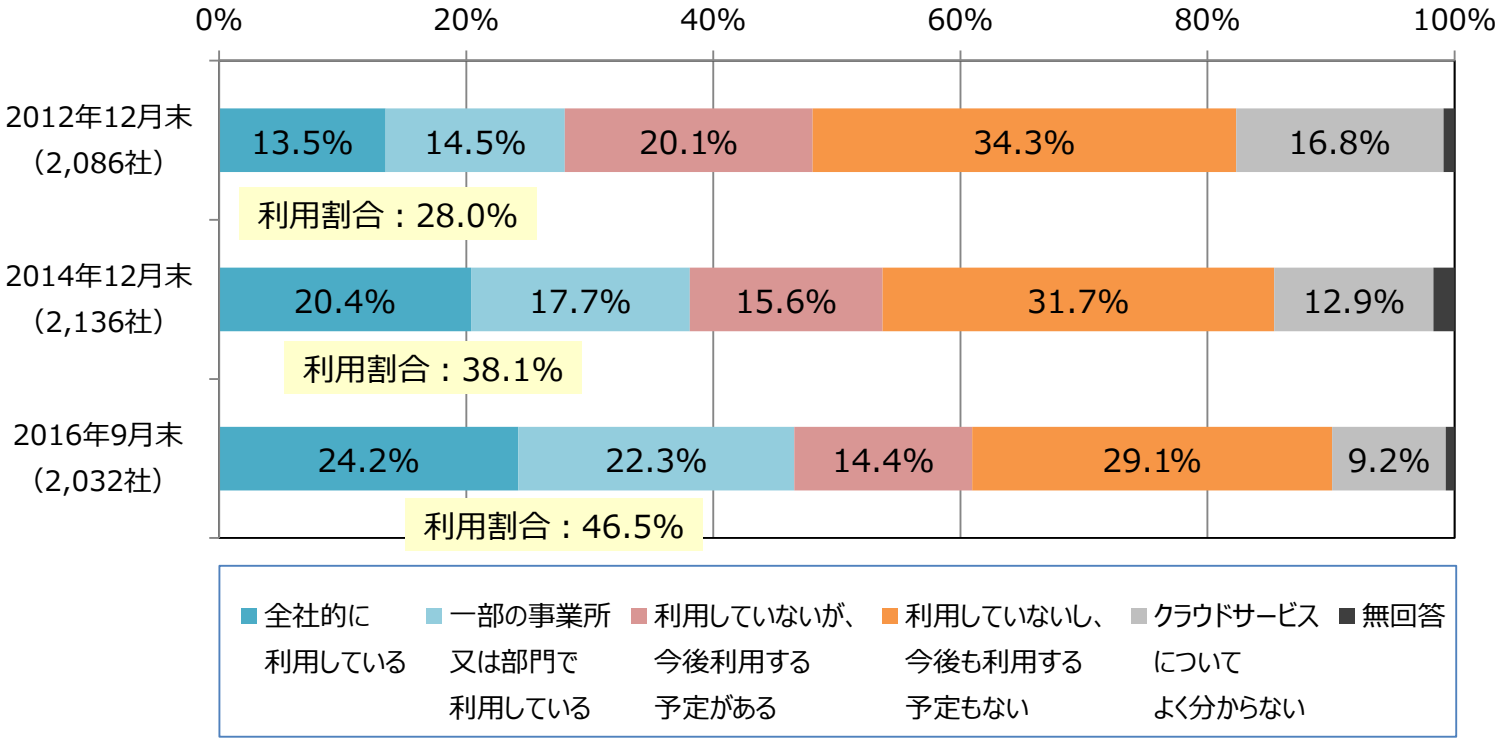
【出所】平成28年通信利用動向調査（企業調査）の調査票［総務省］
http://www.soumu.go.jp/johotsusintokei/statistics/statistics05.html

国内企業におけるクラウドの利用割合

◆国内企業におけるクラウドの利用割合は、増加傾向にあります。

- 通信利用動向調査（企業調査）の1年おきの調査結果を確認すると、国内企業におけるクラウドの利用割合が増加しており、2016年9月末では46.5%の企業がクラウドを利用していることが示されています。
- 「クラウドサービスについてよく分からない」と回答する企業は減少傾向にあり、2016年9月末では9.2%となっています。

通信利用動向調査（企業調査）に基づくクラウドの利用状況



【出所】通信利用動向調査（企業調査）［総務省］に基づいて作成
http://www.soumu.go.jp/johotsusintokei/statistics/statistics05.html

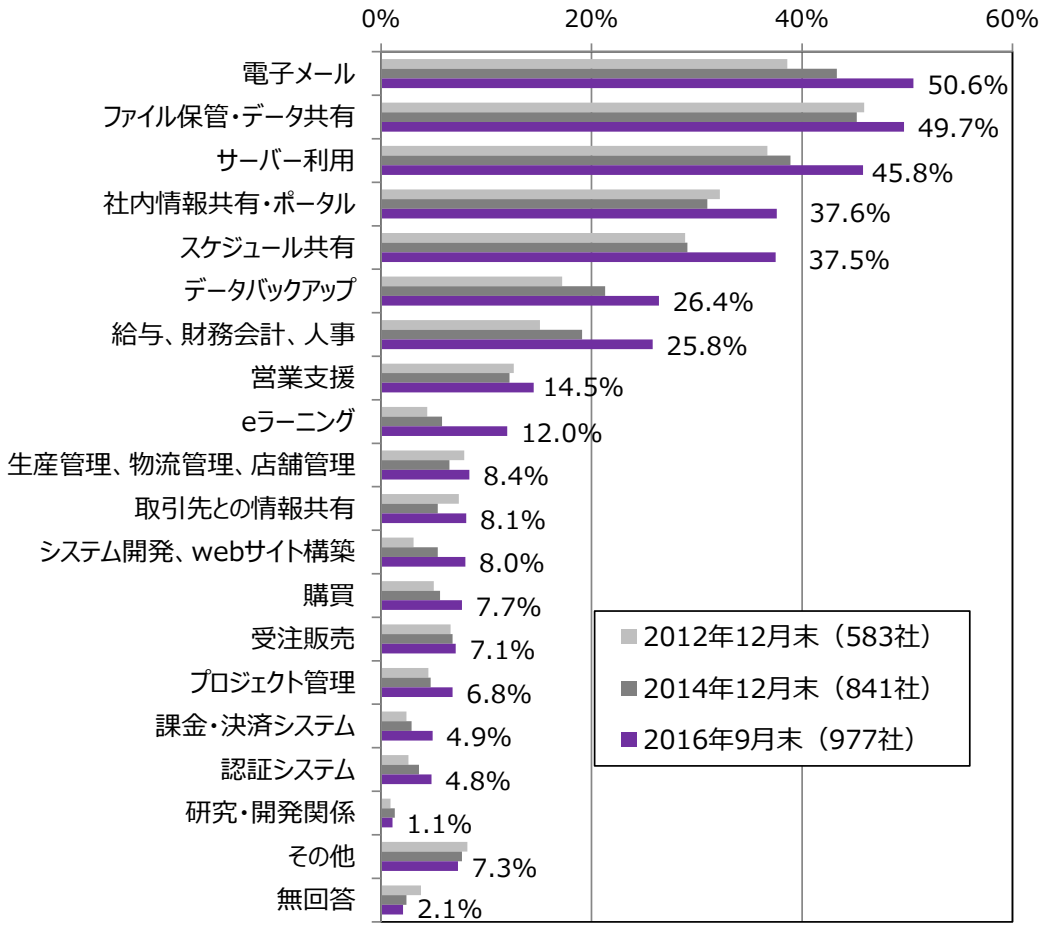
国内企業が利用しているクラウドサービス

◆国内企業が利用しているクラウドサービスとしては「電子メール」が最も多く、様々なサービス種において、利用割合が増加しています。

- 通信利用動向調査（企業調査）の結果から、国内企業が具体的に利用しているクラウドのサービス（複数回答可）を見ると、「電子メール」が最も多く、50.6%を占めています。
  - この質問項目は企業の一部でもクラウドを利用していると回答した企業が母数となった複数回答の質問です。
- 国内企業が利用しているクラウドサービスとして、「電子メール」に次いで「ファイル保管・データ共有」「サーバー利用」「社内情報共有・ポータル」「スケジュール共有」の利用割合が高くなっています。
- これら上位5種のサービス利用割合は2014年12月末から2016年9月末の間でいずれも6%以上伸びています。
  - 2014年12月末から2016年9月末で利用割合が減少したのは「研究・開発関係」と「その他」のみです。

□ 近年、国内企業におけるクラウドサービスの利用は、様々な方面で拡大しています。

具体的に利用しているクラウドのサービス（複数回答）

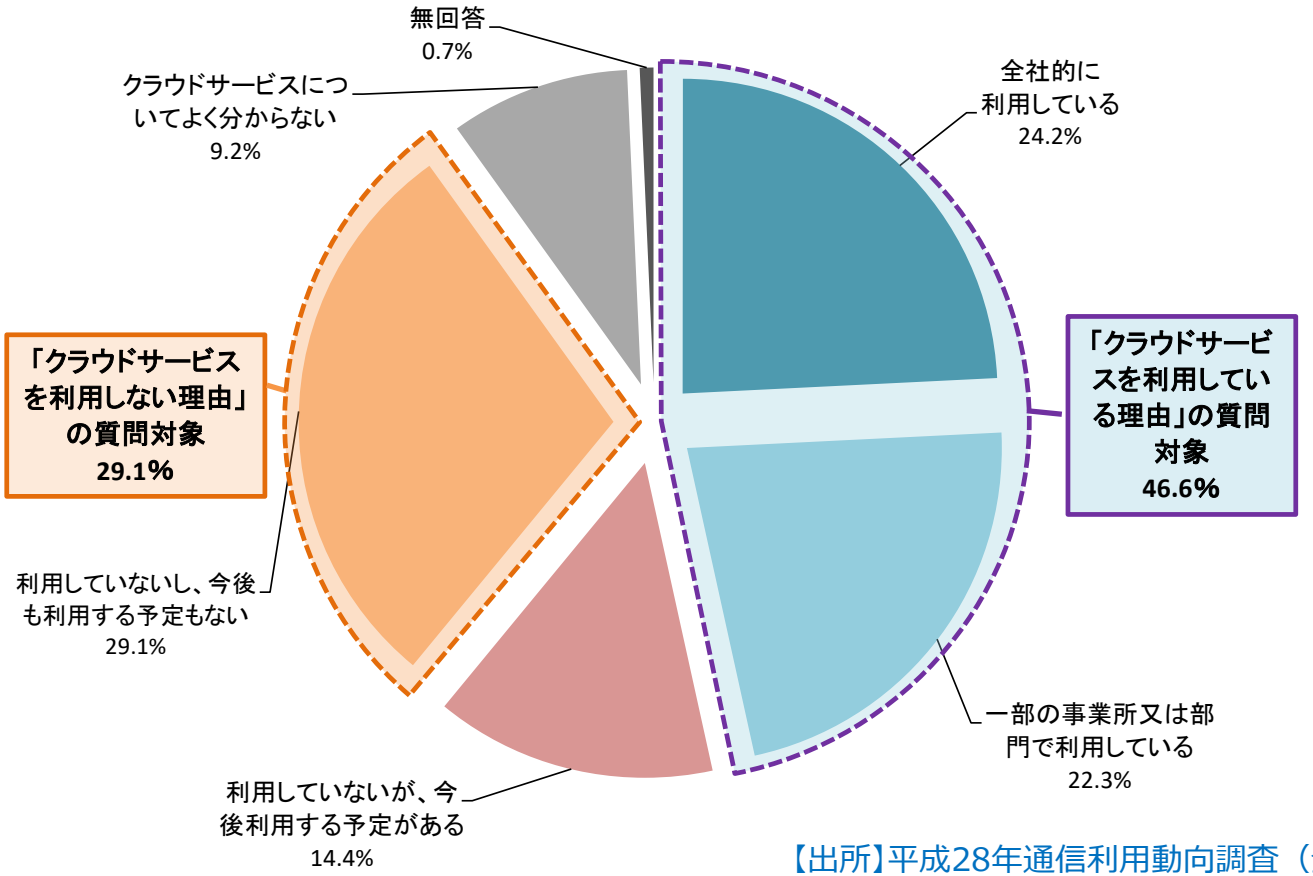


【出所】通信利用動向調査（企業調査）［総務省］に基づき作成
http://www.soumu.go.jp/johotsusintokei/statistics/statistics05.html

［クラウドサービスを利用している/利用しない理由］の回答対象

◆2016年9月末の調査において、クラウドを利用している国内企業の割合は46.6%です。

国内企業におけるクラウドサービスの利用状況（2,032社）[2016年9月末]



【出所】平成28年通信利用動向調査（企業調査）［総務省］

- 通信利用動向調査では、「全社的に利用している」「一部の事業所又は部門で利用している」と回答した企業に利用している理由を尋ねています。
- 調査では「利用していないし、今後も利用する予定もない」と回答した企業には、利用しない理由を尋ねています。

クラウドサービスを利用している理由

◆通信利用動向調査では［クラウドサービスを利用している理由］を調査しています。

- 平成28年通信利用動向調査では［クラウドを利用する理由］を16の回答項目（複数回答可）で尋ねています。

平成28年 通信利用動向調査における［クラウドサービスを利用している理由］の回答項目

番号	回答項目［大分類］	番号	回答項目［大分類］
(01)	既存システムよりもコストが安いから［費用］	(09)	情報漏洩等に対するセキュリティが高くなるから［安全性］
(02)	初期導入コストが安価だったから［費用］	(10)	安定運用、可用性が高くなるから(アベイラビリティ) ［ネット］
(03)	資産、保守体制を社内に持つ必要がないから［費用］	(11)	サービスのラインナップが充実していたから［拡張性］
(04)	システムの容量の変更などが迅速に対応できるから ［拡張性］	(12)	導入スピードが速かったから［拡張性］
(05)	システムの拡張性が高いから(スケーラビリティ) ［拡張性］	(13)	どこでもサービスを利用できるから［ネット］
(06)	サービスの信頼性が高いから［ネット］	(14)	機器を選ばずに同様のサービスを利用できるから ［拡張性］
(07)	システムベンダーに提案されたから ［その他］	(15)	いつでも利用停止できるから［拡張性］
(08)	ライセンス管理が楽だから［費用］	(16)	その他［その他］

- 本教材では、回答項目を表内の［］内に示す表示で「利用する理由」の大分類を行います。

本教材における［クラウドサービスを利用している理由］の大分類

大分類記号	大分類	上の表内の表記	番号
費	費用・手間	［費用］	(01)(02)(03)(08)
拡	拡張性・カスタマイズ性	［拡張性］	(04)(05)(11)(12)(14)(15)
利	インターネット利用・可用性	［ネット］	(06)(10)(13)
安	セキュリティ・コンプライアンス	［安全性］	(09)
他	その他・理由不明	［その他］	(07)(16)

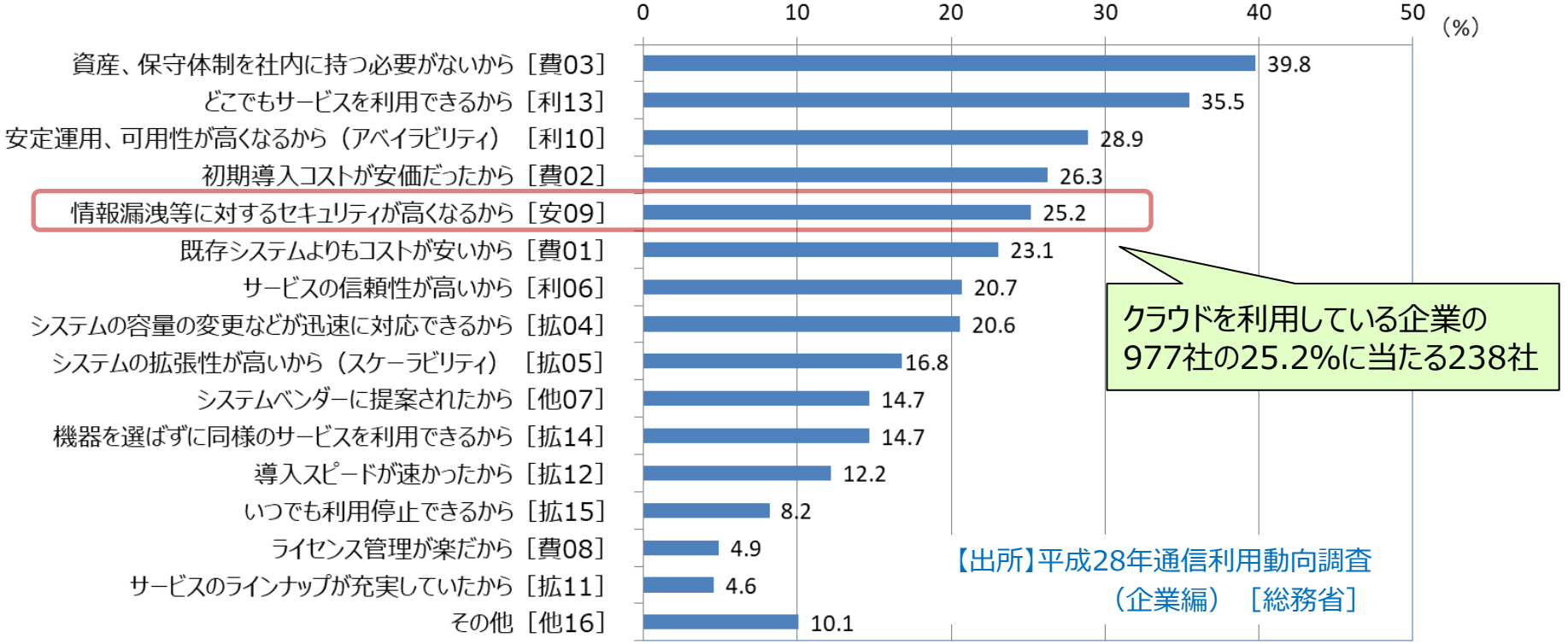
- 可用性（アベイラビリティ）とは、利用ができる機会の多さを指しています。サービスの利用において、利用時間、利用場所、利用端末などの制約が少ない場合を「可用性が高い」と言います。

クラウドサービスを利用している理由の回答割合

◆［利用している理由］として費用やインターネット利用に関する回答割合が高くなっています。

- ［クラウドサービスを利用している理由］として費用に関連する「資産、保守体制を社内に持つ必要がないから」、インターネット利用に関連する「どこでもサービスを利用できるから」がそれぞれ1位、2位の回答割合となっています。
- 回答割合の第4位は「情報漏洩等に対するセキュリティが高くなるから」が25.2%となっています。
  - クラウドの方がそれ以外に比べて、セキュリティが高いと考える割合は、クラウドを利用している国内企業の4分の1を占めています。

［クラウドサービスを利用している理由］の回答割合　（977社：複数回答可）［2016年9月末］



□ クラウドの特徴である「拡張性」よりも「費用・手間」や「インターネット利用」に関する回答が上位に挙がっています。

クラウドサービスを利用しない理由

◆通信利用動向調査では［クラウドサービスをしない理由］を調査しています。

- 平成28年 通信利用動向調査では、クラウドサービスに関して「利用していないし、今後も利用する予定もない」と回答した国内企業に対して［クラウドサービスを利用しない理由］を複数回答可の10の回答項目で尋ねています。

平成28年 通信利用動向調査における［クラウドサービスを利用しない理由］の選択肢

番号	回答内容［大分類］	番号	回答内容［大分類］
(01)	クラウドの導入に伴う既存システムの改修コストが大きい［費用］	(06)	情報漏洩などセキュリティに不安がある ［安全性］
(02)	クラウドの導入によって自社コンプライアンスに支障をきたす［安全性］	(07)	法制度が整っていない ［安全性］
(03)	通信費用がかさむ ［費用］	(08)	必要がない ［その他］
(04)	ニーズに応じたアプリケーションのカスタマイズができない ［拡張性］	(09)	メリットが分からない、判断できない ［その他］
(05)	ネットワークの安定性に対する不安がある ［ネット］	(10)	その他［その他］

- 本教材では、各選択肢を表内の［］内に示す表示で「利用しない理由」の大分類を行います。

本教材における［クラウドサービスを利用しない理由］の大分類

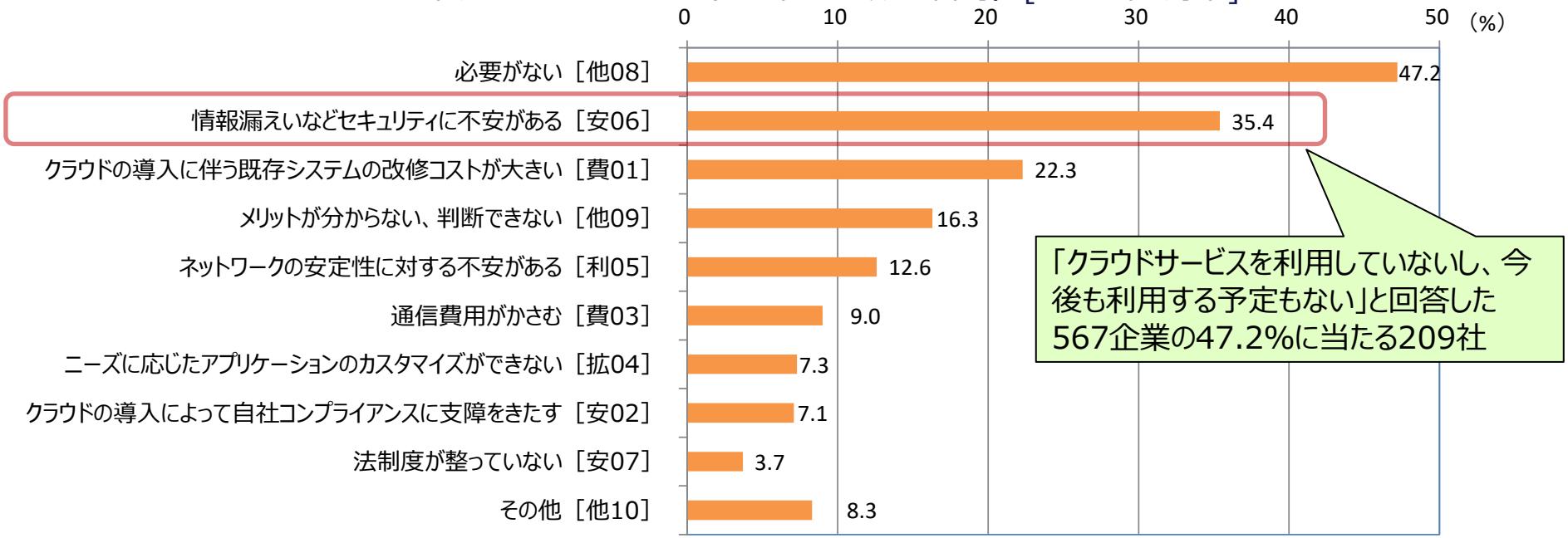
大分類記号	大分類	上の表内の表記	番号
費	費用・手間	［費用］	(01)(03)
拡	拡張性・カスタマイズ性	［拡張性］	(04)
利	インターネット利用・可用性	［ネット］	(05)
安	セキュリティ・コンプライアンス	［安全性］	(02)(06)(07)
他	その他・理由不明	［その他］	(08)(09)(10)

クラウドサービスを利用しない理由の回答割合

◆［利用しない理由］として「必要がない」に次いで、セキュリティ不安が挙がっています。

- ［クラウドサービスを利用しない理由］の回答割合の第1位は「必要がない」の47.2%となっています。
- ［クラウドサービスを利用しない理由］の回答割合の第2位は「情報漏えいなどセキュリティに不安がある」の35.4%となっています。
  - 「情報漏えいなどセキュリティに不安がある」の回答割合の35.4%は、2スライド前に示した「情報漏洩等に対するセキュリティが高くなるから」の25.2%より高くなっていますが、理由に関する調査対象となった企業数が異なるため、回答した企業数は後者の方が多くなっています。

クラウドサービスを利用していない理由　（567社：複数回答可）[2016年9月末]



【出所】平成28年通信利用動向調査（企業編）［総務省］

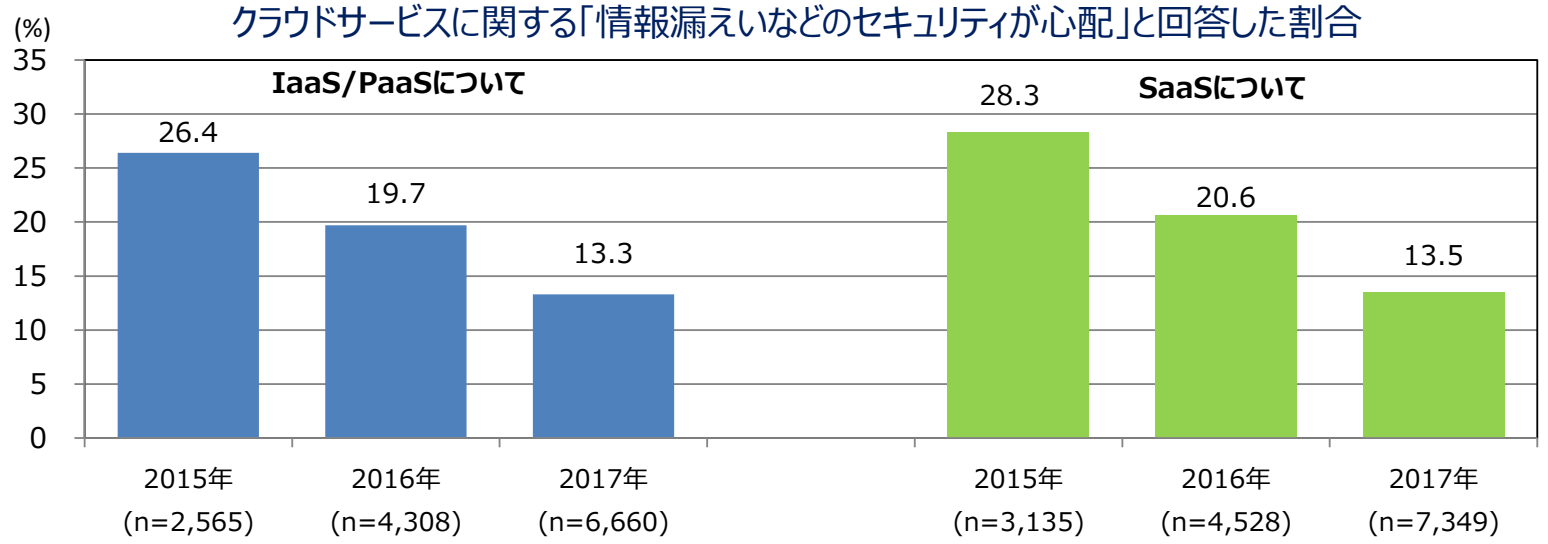
- ［クラウドサービスを利用している理由］においても、［クラウドサービスを利用しない理由］においても、「セキュリティ」に関する回答割合はそれぞれの上位に現れています。

クラウドサービスの情報セキュリティに関する国内企業の意識

◆国内企業におけるクラウドサービスの情報セキュリティに関する懸念は、年々低下しています。

- クラウドサービスに関して、「セキュリティが高くなる」という意見もあれば、「セキュリティが不安」という意見もあります。
  - 平成28年通信利用動向調査（企業調査）においても、［クラウドサービスを利用している理由］と［クラウドサービスを利用しない理由］の両方で上位に挙がっていたのが「セキュリティ」に関する回答でした。
- 株式会社MM総研の調査によれば、クラウドサービスにおける「情報漏えいなどのセキュリティが心配」と回答した割合は年々減少しています。
  - 株式会社MM総研では、企業のネットワーク運用担当者等にウェブアンケートを実施して、クラウドのセキュリティに関する意識を調査しました。
  - クラウドのサービスモデル別に見ると、「IaaS/PaaS」の方が、「SaaS」に比べてセキュリティの心配する回答者の割合が、わずかに低くなっています。
- クラウドサービスのセキュリティの心配する割合が低下した要因として、「クラウド事業者のセキュリティ対策の向上」や「クラウドサービス普及による利用者の理解の進展」などが考えられます。

クラウドサービスに関する「情報漏えいなどのセキュリティが心配」と回答した割合



【出所】株式会社MM総研の2017年12月27日ニュースリリースに基づき作成
https://www.m2ri.jp/news/detail.html?id=279

情報セキュリティの重要性（情報セキュリティに関する事件）

◆コンピュータネットワークにおいて、情報セキュリティは非常に重要です。

- クラウドサービスを含め、コンピュータネットワークにおいて情報セキュリティは重要であり、情報漏洩などの事件があった場合、企業価値や信用を大きく損なうことになります。

情報セキュリティに関する事件の事例

- 2017年9月、アメリカの大手信用情報会社Equifax（エクイファックス）社は、「不正アクセスによってコンピュータネットワークの脆弱性をつかれ、約1億4300万人の個人情報が流出した」と発表しました。
  - 流出した個人情報は、概ねアメリカ在住者に関するものであり、「アメリカ合衆国の人口の44%に相当する」と報道されました。

Equifax社の情報セキュリティに関する事件の説明と注意喚起のウェブサイト

【出所】2017 Cybersecurity Incident & Important Consumer Information［Equifax Inc.］
https://www.equifaxsecurity2017.com/

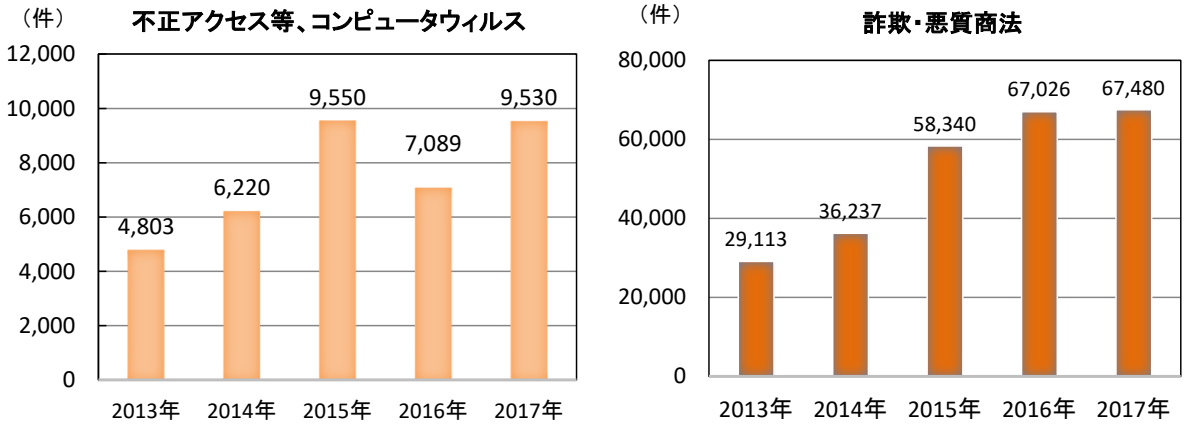
- 大規模な情報セキュリティに関する事件は、当該企業の価値や信用を損なうのみならず、社会問題となり得ます。

情報セキュリティとサイバー犯罪

◆日本において、情報セキュリティに関するサイバー犯罪の相談件数は増加傾向にあります。

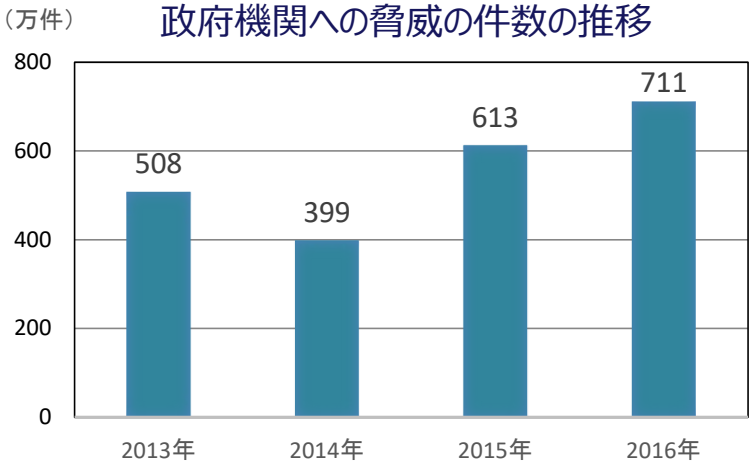
- 都道府県警察に寄せられるサイバー犯罪等の相談件数は、増加傾向にあります。
  - サイバー犯罪の中には、［不正アクセスやコンピュータウィルス］といった情報セキュリティに関するものもあれば、「インターネット上で商品を注文し、代金を振り込んだが、商品が届かない」などの［詐欺・悪質商法］もあります。
- 内閣サイバーセキュリティーセンターの**GSOC（ジーソック）**が、各府省庁等のネットワークに設置したセンサーによって、認知された脅威件数は増加傾向にあります。
  - GSOC（Government Security Operation Coordination team：政府機関情報セキュリティ横断監視・即応調整チーム）は、政府横断的なネットワーク監視、情報セキュリティ上の脅威の検出を行っています。
- 政府組織を含め、価値が高い情報を保有している組織は、サイバー攻撃の対象となりやすい傾向があります。
  - 特定組織が運営するサーバや特定組織の構成員に狙いを絞ったサイバー攻撃は、「標的型攻撃」と呼ばれます。貴重な情報を集めたり、預かっている情報量が大きいサービスは「標的型攻撃」の対象になりやすいと考えられます。

都道府県警察におけるサイバー犯罪等に関する相談件数



【出所】平成29年上半期におけるサイバー空間をめぐる脅威の情勢等について［警察庁］より作成 https://www.npa.go.jp/news/release/2017/20170907001.html

GSOCセンサーで認知された政府機関への脅威の件数の推移



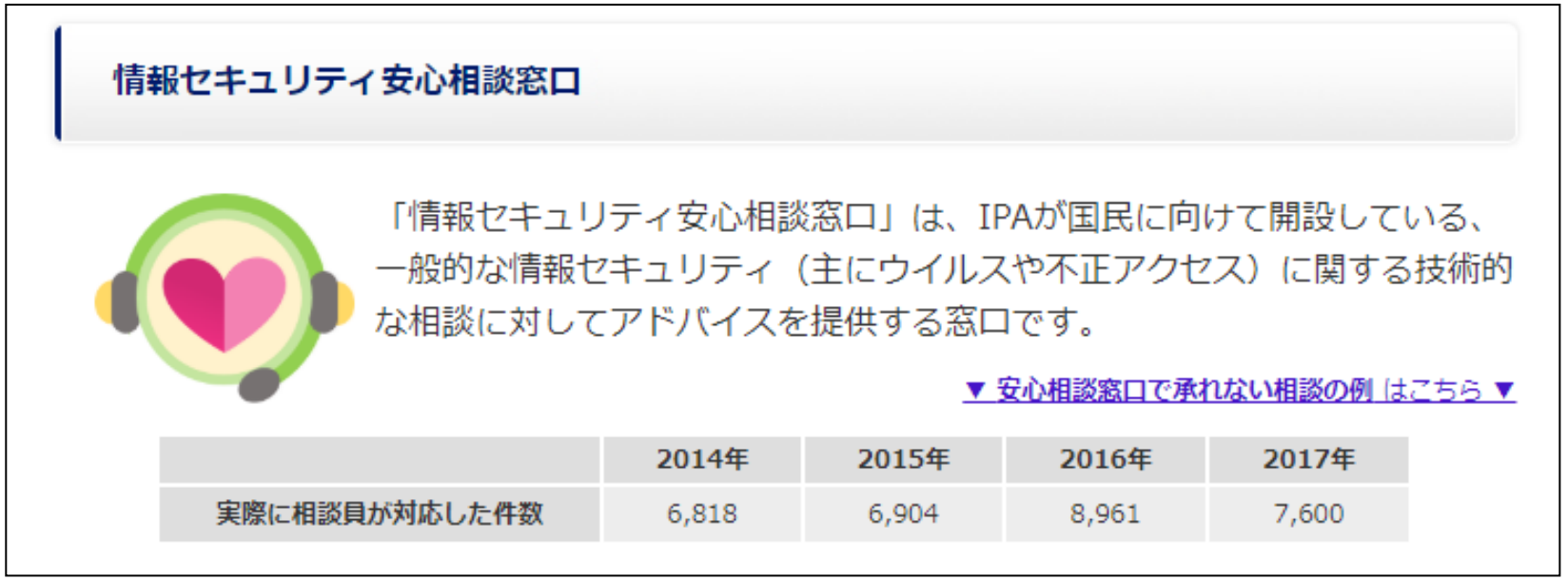
【出所】サイバーセキュリティ政策に係る年次報告［サイバーセキュリティ戦略本部］

情報セキュリティに関する届出・相談窓口

◆日本には情報セキュリティに関する届出・無料相談窓口があります。

- （独）情報処理推進機構（IPA：Information-technology Promotion Agency, Japan）には、**情報セキュリティに関する届出窓口**が設けられています。
  - 届出窓口は「コンピュータウイルス」「不正アクセス」「脆弱性関連情報」に区分され、届け出られた情報は被害の発生や拡大の防止等に利用されます。
- 情報処理推進機構には、**情報セキュリティに関する無料相談窓口**も設けられています。
  - 情報セキュリティ安心相談窓口は、電話・メール・FAX・郵送での相談を受けており、相談料は無料ですが、通話料・通信料は利用者負担です。
  - 「標的型サイバー攻撃」対策として、「標的型サイバー攻撃特別相談窓口」を設けて、一般的な相談とは異なる助言・支援を行っています。

（独）情報処理推進機構が運営する「情報セキュリティ安心相談窓口」のウェブサイト



情報セキュリティ安心相談窓口

「情報セキュリティ安心相談窓口」は、IPAが国民に向けて開設している、一般的な情報セキュリティ（主にウイルスや不正アクセス）に関する技術的な相談に対してアドバイスを提供する窓口です。

▼ 安心相談窓口で承れない相談の例 はこちら ▼

	2014年	2015年	2016年	2017年
実際に相談員が対応した件数	6,818	6,904	8,961	7,600

【出所】情報セキュリティ安心相談窓口［情報処理推進機構（IPA）］
https://www.ipa.go.jp/security/anshin/index.html

情報セキュリティの3つの要素

◆情報セキュリティの3つの要素として「機密性」「完全性」「可用性」が挙げられています。

- 情報セキュリティには、サイバー犯罪への対策を含む「情報漏洩に関する指標」以外にも、天災や機器の故障への対策を含む「正しいデータの保存」に関する指標、「データへのアクセスの可能性」に関する指標があります。
- 日本工業規格（JIS Q 27000:2014）において、**情報セキュリティ**は、情報の**機密性（Confidentiality）**、**完全性（Integrity）**、**可用性（Availability）**の3つを維持することとされています。
  - 情報セキュリティの3つの要素は、その頭文字をとって「情報セキュリティのCIA（シーアイエー）」と呼ばれることもあります。

情報セキュリティの3つの要素とその説明

セキュリティの要素	日本工業規格による用語説明 （ISO/IEC 27000の直訳）	本教材による平易な用語説明
機密性 （Confidentiality）	認可されていない個人、エンティティ又はプロセスに対して、情報を使用させず、また、開示しない特性。	情報漏洩がないこと
完全性 （Integrity）	正確さ及び完全さの特性。	保有する情報が正しく、欠損がないこと
可用性 （Availability）	認可されたエンティティが要求したときに、アクセス及び使用が可能である特性。	利用権限がある者が、いつでも情報を利用できること

- Q 27000：2014によれば「エンティティは，“実体”、“主体”などともいう。情報セキュリティの文脈においては、情報を使用する組織及び人、情報を扱う設備、ソフトウェア及び物理的媒体などを意味する」と記載しています。
- この説明は、情報セキュリティマネジメントシステムの基準と実務を示すために、国際標準化機構 (ISO) と国際電気標準会議 (IEC) 共同で策定したISO/IEC 27000の直訳です。理解しにくい直訳となっているため、第3列には本教材による平易な用語説明を記載しています。

【出所】情報技術－セキュリティ技術－情報セキュリティマネジメントシステム－用語 27000:2014［日本工業規格］

□ 以降では、より平易な言葉で情報セキュリティの各要素を説明します。

情報セキュリティの3要素の説明

◆機密性とは「情報漏洩がないこと」を指しています。

- 機密性の要件として、不正アクセスに対する防御や通信の盗聴防止が挙げられます。
- パスワードなどの認証技術、通信の暗号化技術を利用することで機密性を高めることができます。
  - スマートフォンやUSBメモリでは、紛失や盗難があった場合でも、内部の情報が読みとられないようにすることで「機密性」が高まります。
    （情報セキュリティはインターネットサービスのみならず、情報を蓄積する全ての電子機器を包括する概念です。）

◆完全性とは「保有する情報が正しく、欠損がないこと」を指しています。

- 完全性の要件として、コンピュータウィルスによる情報の消失・改ざんに対する防御が挙げられます。
- セキュリティソフトウェアを導入するなど、情報の消失・改ざんへの対策を行うことで完全性を高めることができます。
- 故障を含め特定の機器のデータが消失しても、バックアップから復旧可能とすることでも完全性を高めることができます。
  - バックアップデータが同じ地域にあると、災害等で同時に利用できなくなるため、異なる国・地域に保存することで完全性を一層高めるケースもあります。

◆可用性とは「利用権限がある者が、いつでも情報を利用できること」を指しています。

- 可用性の要件として、地震などの天災、それに伴う停電への対応策ができていることが挙げられます。
  - 無停電電源装置と呼ばれる短時間の停電に反応して給電を行う機器を取り付けたり、サーバが設置されているデータセンターに自家発電装置を設置したりすることで可用性を高めるケースもあります。
- 機器のメンテナンスや交換に伴う動作不能時間を短縮することで、可用性を高めることができます。
  - 講座2-1に示したように、クラウドにはスケーラビリティ（拡張性）があるため、一般に機器の交換に伴う動作不能時間はありません。

クラウドサービスに関する情報セキュリティの確認項目

◆クラウドサービスは、情報セキュリティ対策を確認した上で選定・利用するべきです。

- **総務省「国民のための情報セキュリティサイト」**では、「クラウドサービス事業者が行っているか」を確認すべき項目として下記の7項目が示されています。
  - 事業者が、これらの情報セキュリティ対策を行うことで「機密性」「完全性」「可用性」を高めることができます。

クラウドサービスを利用する際の情報セキュリティ対策の項目

クラウドサービス事業者が行うべき主要な情報セキュリティ対策
データセンターの物理的な情報セキュリティ対策（災害対策や侵入対策など）
データのバックアップ
ハードウェア機器の障害対策
仮想サーバなどのホスト側のOS、ソフトウェア、アプリケーションにおける脆弱性の判定と対策
不正アクセスの防止
アクセスログの管理
通信の暗号化

【出所】クラウドサービスを利用する際の情報セキュリティ対策［総務省｜国民のための情報セキュリティ対策］
http://www.soumu.go.jp/main_sosiki/joho_tsusin/security/business/admin/15.html

- クラウドサービスは、事業者が上記のような情報セキュリティ対策を継続して適切に行っているかどうかを確認した上で選定・利用するべきです。

□ 実態としては、クラウド事業者がセキュリティを高める観点から情報を一般公開していないケース、情報の理解には専門知識が必要となるケースがあり、各事業者・サービスの情報セキュリティ対策の確認が困難な場合もあります。

クラウドに関する情報セキュリティの第三者認証

◆情報セキュリティ対策には、国際規格としての第三者認証があります。

- 企業やサービスの情報セキュリティ対策のレベルを簡便に確認する方法として、**第三者認証**が挙げられます。
  - 第三者認証とは、当該組織と直接の利害関係がない第三者の審査によって行われる認証を指しています。
- 情報セキュリティに関する代表的な**国際規格として、ISO/IEC 27000シリーズ**が挙げられます。
  - ISO（国際標準化機構）とIEC（国際電気標準会議）が共同で定めた規格であるため、ISO/IECと呼ばれます。
- 従来から一般的な情報セキュリティに関する規格の「ISO/IEC 27001」はありましたが、2015年にクラウドサービスの情報セキュリティに関する規格「ISO/IEC 27017」に関する認証が開始されました。
  - 「ISO/IEC 27001」は一般的な情報セキュリティの構築・維持の基準となる「手順や方法」が記載されており、その基準を満たし審査に合格すれば、「ISMS（Information Security Management System：情報セキュリティマネジメントシステム）認証（取得）」と呼ばれます。
- 3大クラウドサービスと呼ばれるAWS（Amazon）、GCP（Google）、Azure（Microsoft）は、いずれもクラウドサービスのセキュリティに関する規格の「ISO/IEC 27017」の認証を取得しています。
  - ISO/IEC 27017では、組織単位の認定証の中に認証に含まれるクラウドサービスが列挙されています。

ISO/IEC 27017（クラウドサービスセキュリティ）

【出所】ISO/IEC 27017［クラウドサービスセキュリティ］［日本品質保証機構］
https://www.jqa.jp/service_list/management/service/iso27017/

クラウドにおける責任共有モデル

◆「責任共有モデル」は、クラウド事業者と利用者の管理権限に応じた責任分担の考え方です。

- AWSなどのIaaSに関するクラウド事業者は、情報セキュリティに関して一般に**責任共有モデル**を採用しています。
  - IaaSをはじめとするクラウドのサービスモデルに関しては、講座2-2にて紹介しました。
  - 英語の「shared responsibility model」の日本語訳として「責任共有モデル」が定着していますが、「shared」は「共有」よりも「分担」と訳す方が意味が明瞭になります。

AWSにおける責任共有モデルの表記

【出所】責任共有モデル［Amazon Web Services, Inc.］
https://aws.amazon.com/jp/compliance/shared-responsibility-model/

- AWSのウェブサイトにおいて、責任共有モデルの説明として「AWS の責任は"クラウド**の**セキュリティ"」としている一方で、「お客様の責任は"クラウドに**おける**セキュリティ"」としています。
  - クラウドの基礎部分のセキュリティはAWSの責任である一方で、クラウドの内部のセキュリティは顧客（利用者）の責任であることを指しています。

- IaaSでは、利用者がインストールしたOSやソフトウェアに起因するセキュリティのトラブルは、管理権限のある利用者（顧客）の責任となります。
  - IaaSにおいては、利用者は自由に仮想サーバにインストールするOSやソフトウェアを選択できるため、クラウド事業者は仮想サーバの中を管理できません。
  - 情報セキュリティの国際規格で認証されたクラウド事業者であっても、管理権限がない領域に関する情報セキュリティには責任を持つことはできません。

クラウドプラットフォームにおける責任分担に関する表記

◆国際的なクラウド事業者は、責任共有モデルに対応する責任分担を表明しています。

- AWSに限らず、国際的なクラウド事業者は責任共有モデルに対応する責任分担を表明しています。
- GoogleのGCPにおいては、「アプリケーション レベルでのデータ制御」は利用者（顧客）の責任」であることを示しています。
- MicrosoftのAzureでは、クラウドのサービスモデル別・サービス運営の内容別にクラウド事業者と利用者（顧客）の責任分担を示しています。

GCPにおける責任共有モデルに対応する説明（抜粋）

Google Cloud Platform プロジェクトの安全性確保

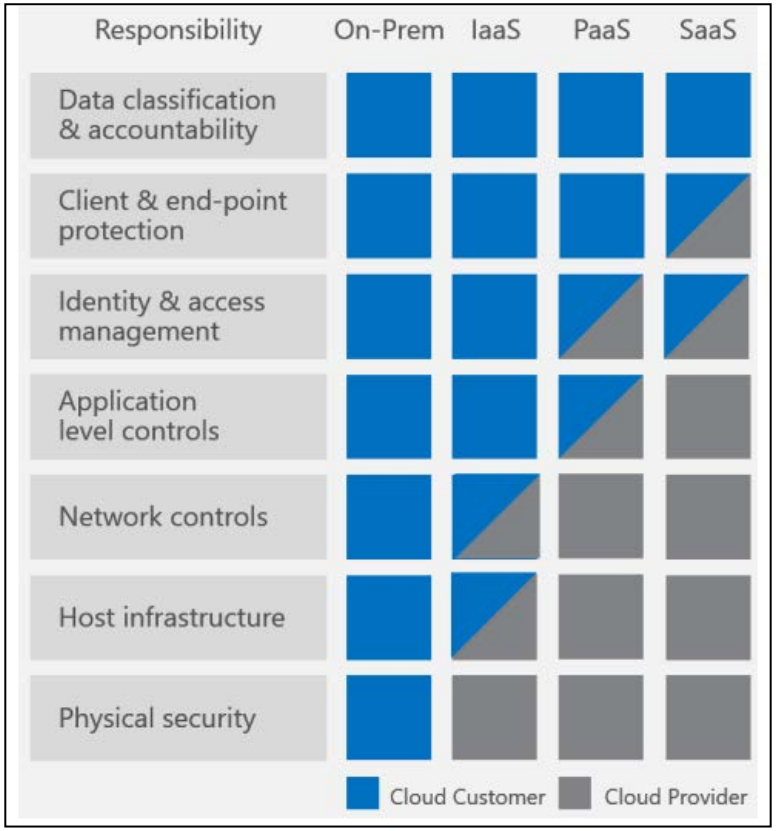
Google はお客様のプロジェクトの安全性を一部担いますが、セキュリティに対する責任は Google 単体ですべてを担えるものではなく、お客様の協力が不可欠です。

機密データの管理

データの重要性はその性質により異なります。Google Cloud Platform では、安全なアプリケーションの構築に必要な基本機能を提供していますが、これらのデータの適切な移動やアクセスをアプリケーションレベルで制御するのはお客様の責任です。これには、エンドユーザーが企業ネットワークやパブリック クラウド インフラストラクチャの外で重要な情報を共有しないよう防ぐ対策（データ損失防止）も含まれます。

【出所】GOOGLE CLOUD PLATFORM のセキュリティ［Google］
https://cloud.google.com/security/?hl=ja

Microsoftによるサービスモデル別の責任分担のイメージ



【出所】Shared Responsibilities for Cloud Computing［Microsoft］https://gallery.technet.microsoft.com/Shared-Responsibilities-81d0ff91

- クラウドに関するセキュリティの責任分担は、サービスモデルのIaaS、PaaS、SaaSの分類にも依存しています。

各サービスモデルにおける情報セキュリティの分担

◆クラウド事業者と利用者の情報セキュリティの責任分担は、サービスモデルによって異なります。

- 2011年にNIST（米国国立標準技術研究所）の研究者は、クラウドサービスに関する情報セキュリティのガイドラインを示した「Guidelines on Security and Privacy in Public Cloud Computing」を公表しました。
- この資料において、クラウドの3種のサービスモデルの情報セキュリティ分担に関して、下表のように記載しています。

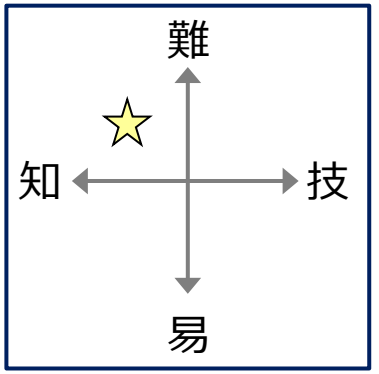
各サービスモデルに関する情報セキュリティの記述

サービスモデル	管理権限と情報セキュリティに関連する説明（抜粋）
IaaS	クラウドの利用者は、一般に搭載するOSや開発環境の選択に関して、高い自由度を持っている。**クラウドの基礎部分を越えるセキュリティ対策は、主としてクラウド利用者が実施する。**
PaaS	クラウドの利用者は、プラットフォーム上のアプリケーションやアプリケーションの環境を管理・設定できる。**セキュリティ対策は、クラウド事業者とクラウド利用者で分割される。**
SaaS	**セキュリティ対策は、主としてクラウド事業者が実施する。**一部を除いて、クラウド利用者はクラウドの基礎部分や個別のアプリケーションの設定を管理、操作できない。

【出所】Guidelines on Security and Privacy in Public Cloud Computing［NIST］から翻訳
https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-144.pdf

- IaaSやPaaSは利用者も情報セキュリティに関する責任や保守作業の一部を分担する一方で、SaaSにおける情報セキュリティ対策は、原則としてクラウド事業者が実施します。
  - IaaSやPaaSでは、仮想サーバ内のOS/ソフトウェアのアップデートといった保守作業は利用者の責任で行う必要があります。
  - SaaSでは、利用者に情報セキュリティに関する専門知識や保守作業を要求しませんが、ログインIDやパスワードの管理は利用者の責任となります。
- 一般にクラウドサービスにおける情報セキュリティは、サービスモデルに応じて利用者にも責任や保守作業の一部分担が求められます。

総務省　ICTスキル総合習得教材



［コース2］データ蓄積

2-4：個人情報の保護と匿名データの利活用

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_2_4.pdf

	1	2	3	4	5
［コース1］データ収集					
［コース2］データ蓄積				●	
［コース3］データ分析					
［コース4］データ利活用					

本講座の学習内容［2-4：個人情報の保護と匿名データの利活用］

【講座概要】

- 2017年に改正施行された個人情報保護法の改正背景、法律の要点を紹介します。
- 改正された個人情報保護法における「個人情報」「個人データ」の定義を説明します。
- 個人情報保護法に基づく個人情報、個人データの取り扱い上の義務・注意点を説明します。
- 法改正によって、新たに加わった「匿名加工情報」の制度概要と利用イメージを紹介します。

【講座構成】

座学

- [1] 個人情報保護法の改正の概要
- [2] 個人情報・個人データベース等・個人データ
- [3] 改正個人情報保護法における義務・規制
- [4] 匿名加工情報の利活用

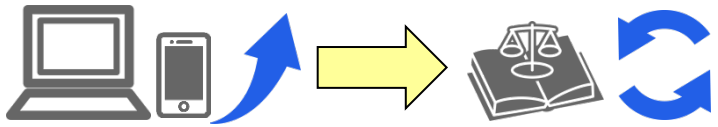
【学習のゴール】

- ✓ 2017年に改正施行された個人情報保護法の改正背景、改正の要点を理解する。
- ✓ 個人情報保護法に基づく個人情報、個人データの取り扱い上の義務・注意点を把握する。
- ✓ 新設された匿名加工情報の制度概要と利用イメージを紹介できる。

個人情報保護法の改正と個人に関するデータの利活用

◆本講座は改正された個人情報保護法に基づき、個人に関する情報の取り扱いを説明します。

- 2015年9月に**個人情報保護法（個人情報の保護に関する法律）**の改正が成立し、2017年5月に施行しました。1
  - 個人情報保護法の正式名称は「個人情報の保護に関する法律」ですが、以降において本教材では通称として普及している「個人情報保護法」と記します。
- 改正の背景には、IoTの普及や個人に関する情報の適正な利活用環境の整備が挙げられます。
  - 改正前の個人情報保護法は2003年の成立ですが、その後、ICTに関する環境や社会情勢は大きく変わってきました。



【出所】個人情報の保護に関する基本方針（個人情報保護委員会）
https://www.ppc.go.jp/files/pdf/290530_personal_basicpolicy.pdf

個人情報保護法の改正背景

- 顔の画像データや指紋データなど、個人情報に該当するかが不明瞭な「グレーゾーン」が拡大してきました。
- インターネットは国境を意識させず、グローバルなデータ流通が進展しました。
- 高度情報通信社会の進展に伴い、個人に関するデータを適正に利活用ができる環境整備が必要となりました。
- 2016年に運用開始されたマイナンバー（個人番号）制度等との関係の整理、明文化が必要となりました。

- 改正された個人情報保護法では、**匿名加工情報**の条項を新設し、個人に関するデータの利活用環境の整備にも貢献しています。
  - 平成29年度の情報通信白書では、「匿名加工情報」の項目新設に関して、「事業者間におけるデータ取引やデータ連携を含むパーソナルデータの利活用促進を目的としたものであり、新事業や新サービスの創出、ひいては、国民生活の利便性の向上につながることが期待される」としています。

【出所】情報通信白書 平成29年度（総務省）
http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/html/nc122120.html

□ 本講座は、ICTの技術面に注目する他の講座と異なり、法務面を中心に個人に関するデータの取り扱いの注意点、活用方法を説明します。

個人情報保護法の第1条（目的）

◆個人情報保護法は、「個人情報の有用性」に配慮した上で「個人の権利利益の保護」を目的とする法律です。

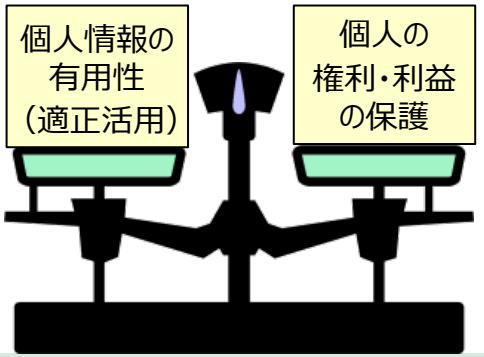
- 改正によって個人情報保護法の第1条（目的）に、**「高度情報通信社会の進展」「個人情報の適正かつ効果的な活用が新たな産業の創出」**という記載が盛り込まれました。
  - 情報通信技術の発達・普及が個人情報保護法の改正背景の一つであることが、第1条にも表れています。
  - 下記の条文紹介における青い太字は、本教材における強調部分を示したもので、個人情報保護法の条文には青い太字での表記はありません。

個人情報保護法　第1条（目的）

この法律は、**高度情報通信社会の進展に伴い個人情報の利用が著しく拡大**していることに鑑み、個人情報の適正な取扱いに関し、基本理念及び政府による基本方針の作成その他の個人情報の保護に関する施策の基本となる事項を定め、国及び地方公共団体の責務等を明らかにするとともに、個人情報を取り扱う事業者の遵守すべき義務等を定めることにより、**個人情報の適正かつ効果的な活用が新たな産業の創出**並びに活力ある経済社会及び豊かな国民生活の実現に資するものであることその他の**個人情報の有用性に配慮しつつ、個人の権利利益を保護することを目的**とする。

【出所】個人情報の保護に関する法律［e-Gov|総務省］
http://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/detail?lawId=415AC0000000057

- 個人情報保護法の第1条には、**「個人情報の適正かつ効果的な活用」「個人情報の有用性に配慮」**という言葉が含まれており、個人情報の保護のみを目的とした法律ではありません。
- 個人情報保護法の第1条は**「個人情報の有用性に配慮しつつ、個人の権利利益を保護することを目的とする」**と締めくくられ、**個人情報の有用性（適正活用）と個人の権利利益の保護のバランスをとることを目的**としています。



個人情報保護法の構成

◆個人情報保護法は、基本法パートと一般法パートの2種類のパートで構成されています。

- 個人情報保護法は、**1～3章の基本法部分**と**4～7章が一般法部分**で構成されています。
  - 基本法パートでは、目的や定義といった官民問わずに適用される内容に加えて、国や自治体の責務・政策の方針が記載されています。
  - 一般法パートでは、民間事業者が対象となる個人情報の保護に関する義務や罰則に加えて、個人情報の保護や匿名加工情報の活用に関与する個人情報保護委員会に関する条項があります。

個人情報保護法の構成（2種類のパート）

章	主な内容
1～3章 （基本法パート）	基本理念および国や地方公共団体の責務・個人情報保護に関する施策
4～7章 （一般法パート）	民間事業者を対象とする個人情報取扱事業者等の義務、個人情報保護委員会

個人情報保護法の章構成

基本法パート	
第1章（1～3条）	総則
第2章（4～6条）	国及び地方公共団体の責務等
第3章（7～14条）	個人情報の保護に関する施策等

一般法パート	
第4章（15～58条）	個人情報取扱事業者の義務等
第5章（59～74条）	個人情報保護委員会
第6章（75～81条）	雑則
第7章（82～88条）	罰則

- 行政機関・独立行政法人・地方自治体は、個人情報保護法の4～7章の対象ではありませんが、個別の法令によって個人情報の保護が義務づけられています。
  - 国の行政機関には「行政機関個人情報保護法」、独立行政法人には「独立行政法人個人情報保護法」、地方自治体には「個人情報保護条例」が適用されます。
  - 対象範囲が広い法律を一般法と呼ぶのに対して、「行政機関個人情報保護法」など、対象範囲が限定された法律を特別法と呼びます。

【出所】個人情報の保護に関する法律［e-Gov|総務省］
http://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/detail?lawId=415AC0000000057

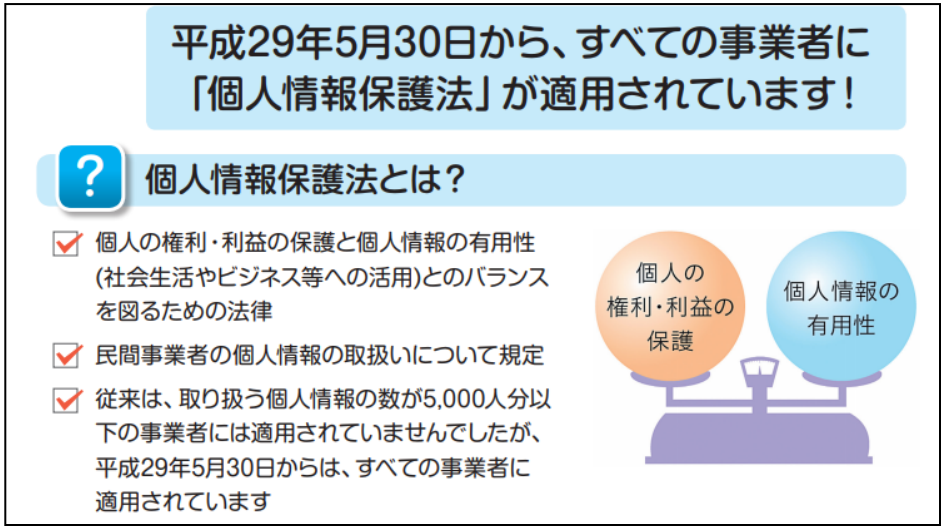
□ 個人情報保護法の内容の説明に先立って、個人情報保護法の対象となる「個人情報取扱事業者」と個人情報保護法の5章に示されている「個人情報保護委員会」を紹介します。

個人情報保護法の対象となる個人情報取扱事業者

◆個人情報保護法では、個人情報取扱事業者を定義し、その義務などを定めています。

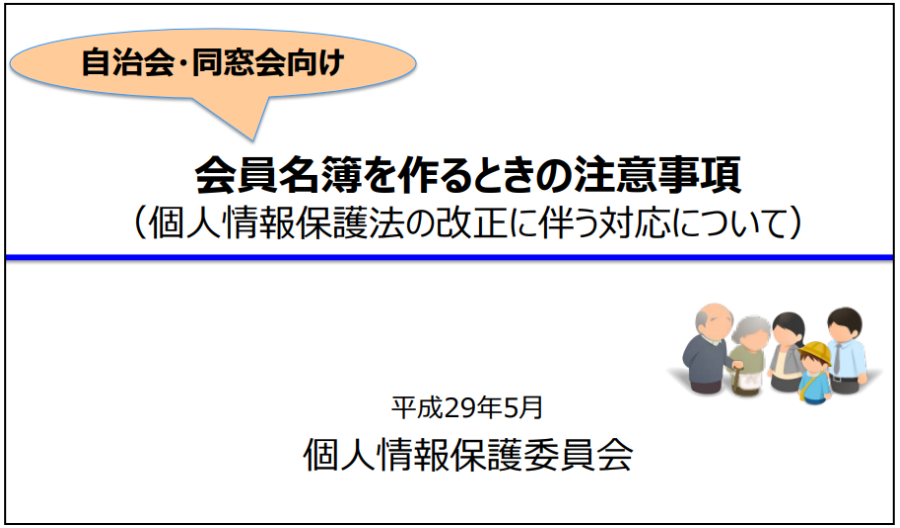
- 個人情報保護法の**対象者**は、個人単位の情報をデータベース化して、**事業の用に供している全ての事業者**です。
  - 個人情報保護法の第2条5に「この法律において「個人情報取扱事業者」とは、個人情報データベース等を事業の用に供している者をいう。」と定められています。それに続き、個人情報取扱事業者の例外として特別法の対象となる国の行政機関、地方公共団体、独立行政法人が挙げられています。
- 2017年5月の個人情報保護法の改正前には、小規模事業者を個人情報取扱事業者の例外とする規定がありましたが、法改正に**よって小規模事業者に対する例外措置が撤廃され、事業者の規模を問わない**ようになりました。
  - 改正前の個人情報保護法では、個人情報取扱事業者の例外として「個人情報によって識別される特定の個人の数の合計が過去6ヶ月以内のいずれの日においても5000を超えない者」と小規模事業者を除外する規定がありましたが、この規定は法改正によって撤廃されました。
- 自治会や同窓会といった小規模の非営利組織も個人情報取扱事業者に該当するようになりました。
  - 個人情報保護法における「事業」は営利・非営利を問わず、対象は法人に限定されずに個人事業主も含まれます。

小規模事業者に関する個人情報保護法の例外措置の撤廃説明



【出所】はじめての個人情報保護法～シンプルレッスン～［個人情報保護委員会］
https://www.ppc.go.jp/files/pdf/1711_simple_lesson.pdf

自治会・同窓会向けの個人情報保護法への対応説明資料



【出所】自治会・同窓会向け会員名簿を作るときの注意事項［個人情報保護委員会］
https://www.ppc.go.jp/files/pdf/meibo_sakusei.pdf

個人情報保護委員会

◆個人情報の保護等を所管する個人情報保護委員会が2016年1月に設立されました。

- **個人情報保護委員会**は、個人情報（マイナンバー（個人番号）を含む。）の有用性に配慮しつつ、その適正な取扱いを確保するために設置された**独立性の高い公的機関**です。
- 個人情報保護委員会は、個人情報保護法および番号法に基づき、それらの関連業務を行っています。
  - 番号法は正式名称は「行政手続における特定の個人を識別するための番号の利用等に関する法律」であり、マイナンバー法と呼ばれることもあります。

【出所】行政手続における特定の個人を識別するための番号の利用等に関する法律［e-Gov|総務省］
http://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/detail?lawId=425AC0000000027

- 個人情報保護委員会の業務として、個人情報に関わる
  - ◆ 特定個人情報の監視・監督に関すること
  - ◆ 苦情あっせん等に関すること
  - ◆ 特定個人情報保護評価に関すること
  - ◆ 個人情報の保護に関する基本方針の策定・推進
  - ◆ 国際協力
  - ◆ 広報・啓発

といった業務を行っています。

- 個人情報保護委員会の業務に挙げられる「特定個人情報」とは「個人番号（マイナンバー）を含む個人情報」を指しています。

個人情報保護委員会のウェブサイト

【出所】個人情報保護委員会　https://www.ppc.go.jp/

個人情報保護委員会が行う広報・啓発と相談対応

◆個人情報保護委員会は、一般向けに広報・啓発や相談対応を行っています。

- 個人情報保護委員会は、個人情報保護法の要点を示したパンフレットを公開するなど、広報・啓発を行っています。
- 個人情報保護委員会では、**個人情報保護法相談ダイヤル**等を設置し、個人情報保護法に関する問い合わせや個人情報に関する苦情などに対応しています。

個人情報保護法に基づく「事業者が守るべき4つのルール」

【出所】はじめての個人情報保護法～シンプルレッスン～［個人情報保護委員会］
https://www.ppc.go.jp/files/pdf/1711_simple_lesson.pdf

個人情報保護委員会の電話相談窓口（2018年3月時点）

個人情報保護法相談ダイヤル

電話番号：03-6457-9849
受付時間：9:30～17:30
（土日祝日及び年末年始を除く）

個人情報保護法の解釈や制度一般に関する疑問や専門的な質問にお答えしたり、個人情報の取扱いに関する苦情をあっせんするため、電話による相談窓口を設置しています。

マイナンバー苦情あっせん相談窓口

電話番号：03-6457-9585
受付時間：9:30～17:30
（土日祝日及び年末年始を除く）

マイナンバー（個人番号）の取扱いに関する苦情の申出についての必要なあっせんを行うため、電話による相談窓口を設置しています。

【出所】個人情報保護委員会　https://www.ppc.go.jp/

□ 本教材では、個人情報保護法の中でも情報通信技術との関わりや匿名加工情報の活用に焦点を当てますが、一般的な説明は個人情報保護委員会の公開資料等を確認し、必要に応じて問い合わせてください。

個人情報の定義（法改正前から継続している定義）

◆個人情報保護法の法改正前から継続している定義として、生存する個人に関する情報で「特定の個人を識別することができるもの」が挙げられます。

- 個人情報保護法における「個人情報」は下記のように定義されており、「一」は法改正前から継続する定義です。

個人情報保護法　第2条（定義）

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って、次の各号のいずれかに該当するものをいう。

一．当該情報に含まれる氏名、生年月日その他の記述等【中略】により特定の**個人を識別することができるもの**

二．個人識別符号が含まれるもの

法改正によって明確に追加された定義

法改正前から継続する定義

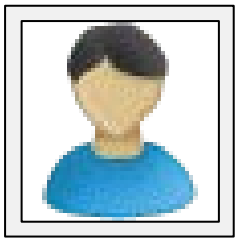
［個人情報の定義の一：特定の個人を識別することができるもの］

◆ **生存する特定の個人が識別**できれば、文字情報の「名刺」、画像の「顔写真」はそれぞれ個人情報となります。

- Eメールアドレスは「yamada-taro@abc.co.jp」と所属企業やフルネームの記載等から個人が識別できれば、個人情報となり得ます。

個人情報保護法第2条1号に該当し、個人情報となる情報

名刺



顔写真

個人情報となり得るEメールアドレス
yamada-taro@abc.co.jp

個人が識別できるEメールアドレス

◆ 複合的な情報から特定の個人を識別できるようであれば、個人情報となり得ます。

- 一般的な氏名であっても、勤務先の組織名や居住地から個人を識別できれば、個人情報となります。

◆ 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ものも個人情報です。

- 「○○株式会社の社長」という情報は、企業ウェブサイト等の情報から容易に照合して、特定の個人を識別することができ、個人情報になり得ます。

個人情報の定義（法改正によって追加された定義）

◆法改正によって、個人識別符号が個人情報の定義に追加されました。

- 法改正によって、**個人識別符号**（生体情報、公的機関が関わるID）が個人情報の定義に追加されました。
  - 従来、個人情報としての取り扱いがグレーゾーンであった生体情報や公的機関が関わるIDが個人情報に含まれることが法律に明記されました。

［個人情報の定義の二：個人識別符号が含まれるもの］

◆ 個人識別符号は大別して2種類があり、個人情報保護委員会の政令に具体的に示されています。

【出所】個人情報の保護に関する法律施行令［個人情報保護委員会］
https://www.ppc.go.jp/files/pdf/290530_personal_cabinetorder.pdf

個人情報の保護に関する法律施行令（第1条 一）

細胞のDNA	歩行の姿勢
顔の画像・動画	手の静脈の形状
瞳の虹彩	指紋・掌紋
声の特徴	

◆「第1条 一」に示される個人識別符号は、**身体の一部の特徴を変換したデータ**が該当します。

- 1号個人識別符号は、ICTの進展によって生体認証に利用することが考えられる情報に対応しています。

細胞のDNA

瞳の虹彩

声の特徴

指紋

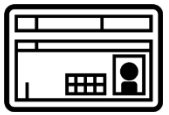
個人情報の保護に関する法律施行令（第1条 二～八）

旅券番号	マイナンバー（個人番号）
基礎年金番号	医療保険の個人番号
運転免許証番号	これらに準ずるものとして、個人情報保護委員会規則に定めるもの
住民票コード	

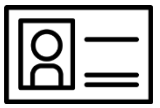
◆「第1条 二～八」に示される個人識別符号は、**公的機関が関わる個人ID**が該当します。

- 一般の個人・事業者は、2号個人識別符号から個人を特定することは困難ですが、個人情報に含まれることが明記されました。

旅券番号



運転免許証番号



マイナンバー（個人番号）

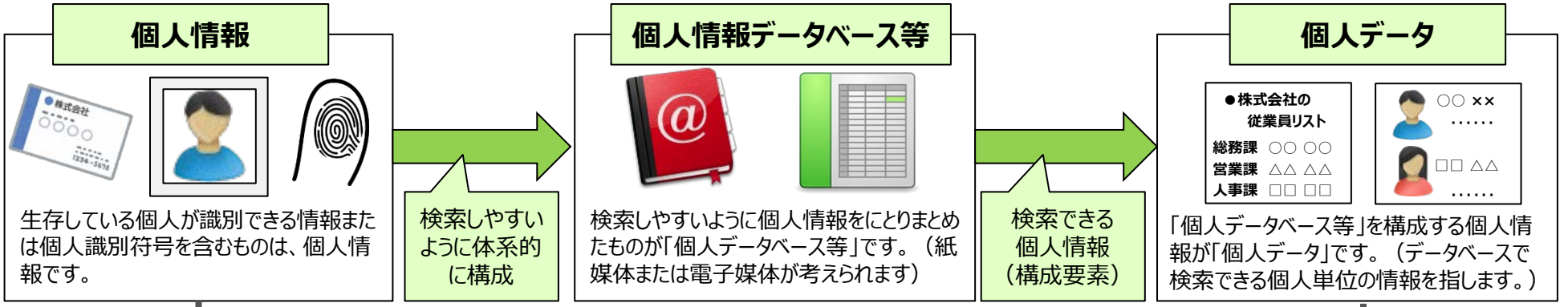
医療保険の個人番号

個人情報データベース等と個人データ

◆個人情報保護法では、「個人情報データベース等」「個人データ」という概念があります。

- 個人情報保護法の第2条4において**個人情報データベース等**とは「個人情報を含む情報の集合物であって、検索することができるように体系的に構成したもの」と定義されています。
  - データベースの定義として「検索ができること」を挙げているのは、講座2-1で示した著作権法と同様です。
- 個人情報保護法の第2条6において**個人データ**とは「個人情報データベース等を構成する個人情報」と定義されています。
  - 個人が識別できる名刺や写真はそれ自体が「個人情報」ですが、「個人データベース」に格納されることで、初めて「個人データ」となります。
- 個人情報保護法において「個人情報」および「個人データ」にはそれぞれ後述する規制が定められています。

個人情報保護法における「個人情報」「個人データベース等」「個人データ」の関係



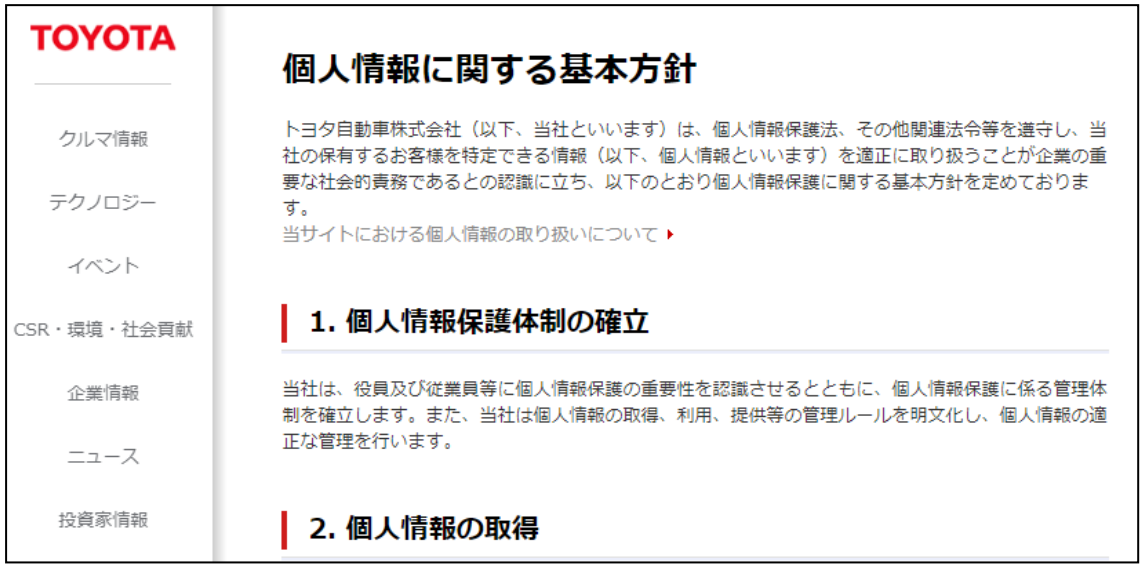
- 個人情報に関しては「利用目的の特定（第15条）」「利用目的による制限（第16条）」「取得に際しての利用目的の通知等（第18条）」が定められています。
- 個人データに関しては「安全管理措置（第20条）」「第三者提供の制限（第23条）」「開示（第28条）」が定められています。

個人情報の利用目的の特定および通知・公表

◆個人情報の取得時には利用目的を特定し、通知・公表することが義務づけられています。

- 事業者は**「個人情報」の取得時に利用目的を特定**し、**公表または本人に通知することが必要**となります。
  - 個人情報保護法の第18条に「個人情報取扱事業者は、個人情報を取得した場合は、あらかじめその利用目的を公表している場合を除き、速やかに、その利用目的を、本人に通知し、又は公表しなければならない。」と定められています。
- 個人情報を取得するたびに通知・公表する必要性を避けるため、ウェブサイトを持つ企業・組織はあらかじめ**「個人情報保護方針」**や**「プライバシーポリシー」**を公表しているケースが一般的です。
  - 「個人情報保護方針」と「プライバシーポリシー」は実質的に同じものを指しており、「個人情報保護方針（プライバシーポリシー）」と括弧書きで併記しているケースもあります。

トヨタ自動車株式会社の「個人情報に関する基本方針」



【出所】個人情報に関する基本方針［トヨタ自動車株式会社］
http://www.toyota.co.jp/jpn/terms/privacy_statement.html

日産自動車株式会社の「プライバシーポリシー」

【出所】プライバシーポリシー［日産自動車株式会社］
https://www3.nissan.co.jp/siteinfo/privacy.html

個人情報の利用目的の制限（利用目的の公表例）

◆取得した個人情報は、本人に通知・公表した利用目的の範囲でのみ利用できます。

- 事業者が取得した個人情報は、原則として**本人に通知・公表した利用目的の範囲内でのみ利用可能**です。
  - 例外として「法令に基づく場合」「人の生命、身体又は財産の保護を目的とし、本人の同意を得ることが困難な場合」等が挙げられます。
- 通知・公表を行う利用目的には、利用可能性のある複数の利用目的を列挙して示すことができます。
  - 共同利用先をあらかじめ本人に通知・公表しておけば、グループ企業や提携企業などとも情報を共有することが可能です。
- 一般的な利用目的を公開しておき、各個人情報の取得時に示した利用目的を追加する旨の記載も可能です。
  - 下記の「左側の事例の5」および「右側の事例の最後の注意書き」が、各情報の取得時に明示した場合の追加に関する記載となっています。

トヨタ自動車株式会社における「利用目的」（2018年3月時点）

1. 当社で取り扱っている商品・サービスなどに関する営業上のご案内を行うこと。
2. 商品の企画・研究開発・品質向上のほか、お客様満足度向上策の策定などのために、お客様にアンケート調査を行うこと。

【中略】

5. その他個人情報取得時に明示した利用目的。
6. 法令の定め又は行政当局の通達・指導などに基づく対応を行うこと。なお、トヨタ販売店を通じてお客様へのご案内を行う場合は、個人情報をトヨタ販売店に提供させていただきます。
7. 当社「お客様相談センター」へご相談いただいた場合、適切なご対応を行うため、必要に応じて以下の個人情報をトヨタ販売店・当社提携会社へ、電話・書面・電子媒体などにより提供すること。ただし、お客様のお申し出により、トヨタ販売店・当社提携会社への提供を停止させていただきます。

＊提供する個人情報の項目：氏名、住所、年齢、電話番号及びお客様のご相談内容に関する情報

【出所】個人情報に関する基本方針［トヨタ自動車株式会社］
http://www.toyota.co.jp/jpn/terms/privacy_statement.html

日産自動車株式会社における「利用目的」（2018年3月時点）

1. 当社が取扱う商品・サービス等について、または各種イベント・キャンペーン等の開催について、宣伝印刷物の送付、電話、電子メールの送信等によりご案内すること。
2. 商品開発またはお客さま満足度向上策等の検討のために、アンケート調査を実施すること。
3. お客さまの個人情報を書面または電子媒体により下記の第三者に提供すること。ただし、お客さまご本人のお申し出があった場合、第三者提供を停止いたします。

提供先：

1）当社の子会社及び関連会社並びに当社の特約販売会社

2）当社と情報提供契約を締結した者

ただし、ウェブサイト、書面等により、別途の利用目的が公表されている場合、または取得に当たり、お客さまに利用目的を明示した場合においては、その利用目的が優先されるものとします。

【出所】プライバシーポリシー［日産自動車株式会社］
https://www3.nissan.co.jp/siteinfo/privacy.html

個人データに関する安全管理措置

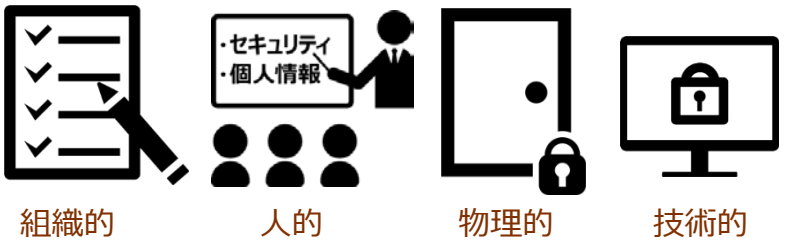
◆個人データの取り扱いには、従業者・委託先を含めて安全管理措置の義務があります。

- 個人情報取扱事業者は、**個人データの取り扱いに安全管理措置の義務**があります。
  - 個人情報保護法の第20条に「個人情報取扱事業者は、その取り扱う個人データの漏えい、滅失又はき損の防止その他の個人データの安全管理のために必要かつ適切な措置を講じなければならない。」と定められています。
- 個人情報保護法のガイドラインにおいて、安全管理措置には**組織的**、**人的**、**物理的**、**技術的**の4種があるとされています。

【出所】個人情報の保護に関する法律についてのガイドライン（通則編）[個人情報保護委員会]
https://www.ppc.go.jp/files/pdf/guidelines01.pdf

4種の安全管理措置

安全管理措置の分類	安全管理措置の主な内容
組織的安全管理措置	組織体制の整備、取扱規程等に基づく運用
人的安全管理措置	事務担当者の監督と教育
物理的安全管理措置	入退館（室）の管理、盗難の防止
技術的安全管理措置	アクセス者の識別と認証、不正アクセスの防止



- 個人情報取扱事業者は、**従業員**および**個人情報を取り扱う委託先**における安全管理の監督義務があります。
  - 安全管理の監督とは、対象に安全管理の状況を報告させたり、対象の安全管理の状況を確認したりして、必要に応じた指導を行うことを指しています。
  - 個人情報取扱事業者の個人情報を委託先で取り扱う業務委託を行う場合は、安全管理の監督が可能となる委託契約を結ぶ必要があります。
  - 「業務の委託」「事業の承継」「本人にあらかじめ通知または容易に知り得る情報とした上での共同利用」は、個人情報の第三者提供には当たりません。

クラウドサービス等への個人データの保存

◆委託契約がなくとも、クラウドサービス等に個人データを預けることができるケースがあります。

- 個人情報取扱事業者がクラウド等のインターネットサービスに個人データを預ける場合は、**「データを預かる者が、個人データを取り扱うことになっているか否か」**で、結ぶべき契約の種類が異なります。

◆ データを預かる者が、個人データを取り扱うことになっている場合

- データを預かる者が、個々の個人データを確認し、それに応じたサービスを行うケースが該当します。

➡ 安全管理の監督に関する取り決めを含む委託契約が必要

サービスの事例
- 個人宛が含まれる郵便物の印刷・送付サービス
- 入力された個人情報に応じた人材マッチングサービス

◆ データを預かる者が、個人データを取り扱うことになっていない場合

- データを預かる者が、契約条項によって個人データを取り扱わない旨を定められているケースや個人データを預かっている認識がなく、データの内容を見ることもないケースが該当します。

➡ 通常のサービス利用契約で、個人データを預けることが可能

サービスの事例
- サービス提供者による入力情報の確認不可が明記されているウェブアンケートフォーム設置サービス
- 普及しているクラウドサービス（仮想サーバや電子メールサービス等）

- 普及している**クラウドサービス（仮想サーバや電子メールサービス等）は、一般に「データを預かる者が、個人データを取り扱うことになっていない場合」に該当**し、**通常のサービス利用契約によって個人データを預けることが可能**です。

- データを預かる者がデータの内容を見ない場合は電子媒体の取り扱いでも、個人データを含む紙媒体を銀行の貸金庫に保管できること、郵送できることと同様です。
- 個人データを含むファイルが「紙媒体であれば、書留郵便で送付する」「電子媒体であれば、パスワードを設定する」といった個人情報取扱事業者自身の安全管理措置は求められます。

【出所】「個人情報の保護に関する法律についてのガイドライン」及び「個人データの漏えい等の事案が発生した場合等の対応について」に関するＱ＆Ａ（平成 29 年5月30日更新版）［個人情報保護委員会］の「Q&Aの5-33、5-34」に基づき作成
https://www.ppc.go.jp/files/pdf/kojouhouQA.pdf

通常のサービス利用契約で
個人データを預けることが可能

普及しているクラウドサービス

個人データの開示請求等への対応（プライバシーマークとの関係）

◆個人情報取扱事業者は、本人からの個人データの開示請求等に対応する義務があります。

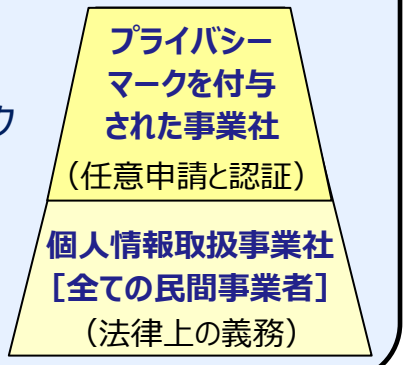
- 個人情報取扱事業者は、本人からの**開示請求に応じて、原則として個人データを開示する義務**があります。
  - 開示しなくても良い例外のケースとして第28条2には「本人又は第三者の生命、身体、財産その他の権利利益を害するおそれがある場合」「本人又は第三者の生命、身体、財産その他の権利利益を害するおそれがある場合」「他の法令に違反することとなる場合」が挙げられます。
  - 開示しない旨を決定した場合および当該個人データが存在していない場合は、当人にその旨を速やかに通知する必要があります。
- 個人情報取扱事業者の保有している**個人データが事実でない場合は、本人は内容の訂正・追加・削除を請求する**ことができ、**個人情報取扱事業者は応じる義務**があります。
- 個人情報を利用目的以外に使った場合、個人情報を**不正に取得した場合は、本人は個人データの利用の停止や消去を請求する**ことができ、**個人情報取扱事業者は原則として応じる義務**があります。

【個人情報保護法と日本情報経済社会推進協会（JIPDEC）のプライバシーマークの関係】

- ◆ 一般財団法人 日本情報経済社会推進協会では、事業者の個人情報を取り扱う仕組みと運用のレベルを評価し、協会で定めた基準を超えたレベルの事業者には、プライバシーマークの使用を認めています。
  - 日本情報経済社会推進協会は、従来の英語名称である「Japan Information Processing and Development Center」を由来として、通称JIPDEC（ジプデック）とも呼ばれます。
- ◆ プライバシーマークを付与された事業者には、個人情報保護法に定められたレベルを超える個人情報の保護・本人への対応が求められます。
- ◆ プライバシーマーク付与事業社に対して、本人は個人データの利用停止や消去を請求でき、規約により当該事業者には無条件で応じる義務があります。
  - 個人情報保護法では、目的外利用や不正取得等があった場合という条件が必要です。

【出所】「個人情報」の開示・削除等と利用停止等を求めるとき［日本情報経済社会推進協会］
https://privacymark.jp/wakaru/kouza/theme4_05.html

日本情報経済社会推進協会のプライバシーマーク



第三者提供に関する本人同意

◆個人情報取扱事業者が個人データを第三者に提供する場合は、本人の同意が必要です。

- 個人情報取扱事業者が個人データを**第三者に提供する場合は、原則として本人の同意が必要**です。
  - 例外として「法令に基づく場合」「人の生命、身体又は財産の保護を目的とし、本人の同意を得ることが困難な場合」等が挙げられます。
  - 第三者提供について、あらかじめ本人の同意を得ていなければ、第三者提供の可否について改めて本人に確認をする必要があります。
- 本人の同意をもって個人データを第三者に提供を行う場合でも、**データの提供者・受領者それぞれに記録・保存の義務**があります。

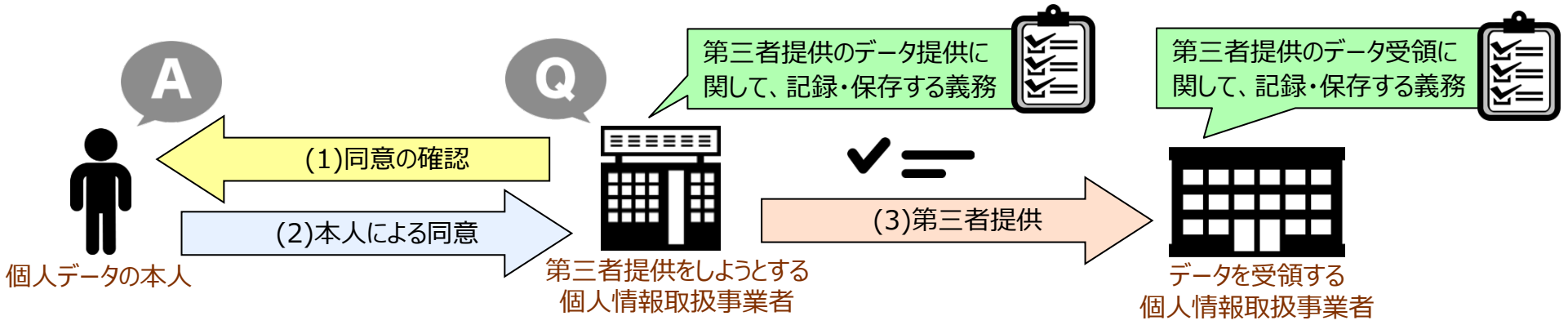
◆データ提供者となる個人情報取扱事業者は、データの受領者、本人の氏名等を記録し、保存する義務があります。

◆データ受領者となる個人情報取扱事業者は、データの提供元、取得経緯、本人の氏名等を記録し、保存する義務があります。

- データの提供側・受領側の保存期間はそれぞれ、本人に対する物品又は役務の提供に関連するデータ提供で契約書等がある場合は1年、それ以外の場合は3年であることが、「個人情報の保護に関する法律施行規則」に定められています。

【出所】個人情報の保護に関する法律施行規則[個人情報保護委員会]
https://www.ppc.go.jp/files/pdf/290530_personal_commissionrules.pdf

本人同意に基づく個人データの第三者提供の流れ



オプトイン・オプトアウトによる本人同意

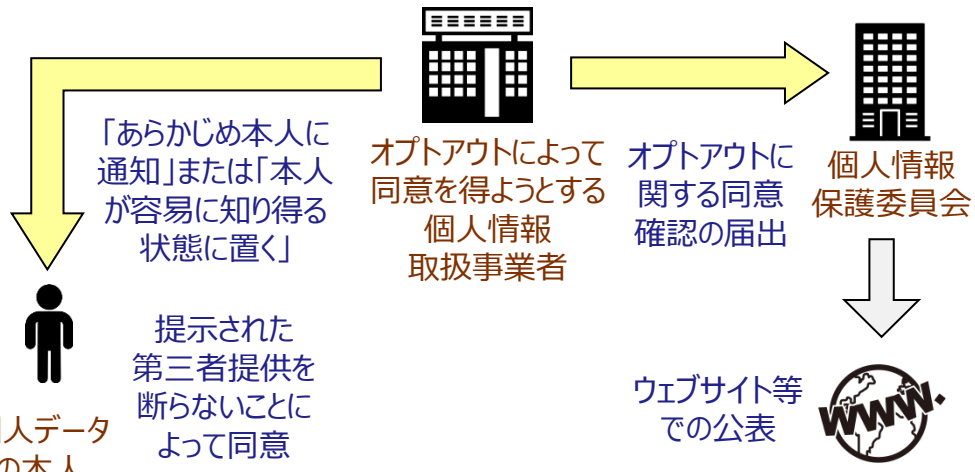
◆第三者提供に関する本人同意の確認方法は、法改正によって厳格化されました。

- 第三者提供に関する本人の同意の確認方法は、**オプトイン**と**オプトアウト**の二種類の形式があります。
  - オプト（opt）には「選ぶ、選択する」という意味があり、オプトインは「同意に入ることを選ぶ」、オプトアウトは「同意に入らないことを選ぶ」ことを指します。

同意の形式	確認内容	無回答の扱い
オプトイン（opt in）	「提示した情報での第三者に提供しても良いですか？」と尋ねて、本人から「提供しても良い」との回答を得る形式	本人からの回答なしは「同意なし」とみなす
オプトアウト（opt out）	「提示した情報での第三者への提供を断るなら教えてください。」と尋ねて、本人から「提供を断る」との回答がなければ、同意と見なす形式	本人からの回答なしは「同意あり」とみなす

- 法改正によって、**オプトアウト**によって本人同意を得ようとする事業者は、「第三者に提供される個人データの項目」「本人の求めを受け付ける方法」等を**個人情報保護委員会に届け出る義務**が課せられました。
  - 届出義務に加えて、第三者への提供に関する内容を「あらかじめ本人に通知」または「本人が容易に知り得る状態に置く」必要があります。
  - オプトアウトによる本人同意の届出を受けた個人情報保護委員会は、個人情報保護法第23条4に基づきウェブサイト等で届出の内容を公表します。

オプトアウトによる本人同意確認の手続きの流れ



個人情報保護委員会ウェブサイトのオプトアウト届出書一覧

【出所】オプトアウトの届出書一覧［個人情報保護委員会］
https://www.ppc.go.jp/personal/preparation/optout/publication_list/

匿名加工情報の条項の新設

◆法改正で、本人同意不要でのデータ提供が可能な匿名加工情報の条項が新設されました。

- 個人情報保護法における**匿名加工情報**は、特定の個人を識別できないように個人情報を加工し、**当該個人情報を復元できないようにした情報**を指しています。

個人情報から匿名加工情報への置き換え例

個人情報

氏名：○○××
（33歳の男性）
10月11日（水）の19時32分に
東京都港区のスーパー□□で
食パンと紅茶とミートボールを購入

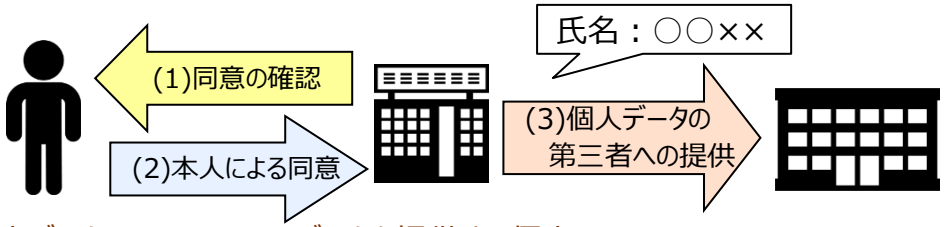
項目	加工の取り扱い
氏名	削除（仮IDに置き換え）
年齢	10歳区分の年齢層に置き換え
性別	加工なし
購入日	月単位の平日・休日に置き換え
購入時間	3時間単位に置き換え
購入区域	都道府県情報
購入店	削除
購入商品	商品のカテゴリに置き換え

匿名加工情報

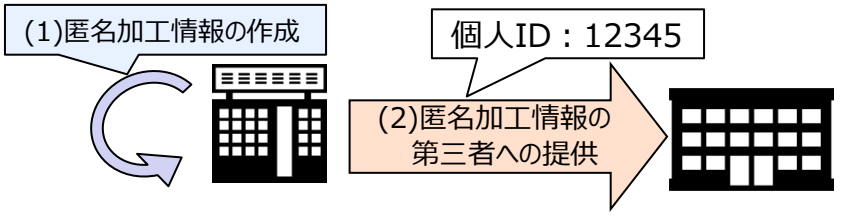
個人ID：12345
（30歳代の男性）
10月平日の18時～21時
東京都のスーパーマーケットで
パン、飲料、惣菜を購入

- **匿名加工情報は**個人データとは異なり、**本人の同意不要で第三者へのデータの提供を行うことが可能**です。
  - 個人情報保護法の第2条10より、匿名加工情報を取り扱う事業者は、匿名加工情報取扱事業者と呼ばれます。

個人データの第三者への提供の流れ



匿名加工情報の第三者への提供の流れ



匿名加工情報の利用事例

◆匿名加工情報は、ビジネスやヘルスケアなど広範な活用が期待できます。

- 匿名加工情報は、本人同意不要で第三者提供ができることから、様々な活用が期待されます。
- 一般財団法人 日本情報経済社会推進協会はビジネスに関する「匿名加工情報の事例集」を公開しています。
  - 日本情報経済社会推進協会は、プライバシーマークの付与機関であるとともに認定個人情報保護団体でもあります。
  - 認定個人情報保護団体とは、「業界・事業分野ごとの民間による個人情報の保護の推進を図るために、自主的な取組を行うことを目的として、個人情報保護委員会の認定を受けた法人」を指しています。

日本情報経済社会推進協会が示した匿名加工情報の利活用の事例

提供主体	提供先	匿名加工の元データ	利用例
自動車整備工場	自動車販売店	• 顧客データ • 車両データ	自動車販売店は、性別や年齢、車両の色等を分析し、自動車のマーケティングに活用
質屋	調査会社	• 顧客データ • 本人確認書類データ	調査会社は、本人確認書類（免許証、保険証、パスポートなど）の利用状況を調査分析
商店街	店舗設置を検討する事業者	• 顧客データ • 購買記録データ	事業者は、顧客層、購買日時、購買金額等を分析し、商店街への店舗設置、提供サービスを検討
交通ナビゲーションサービス企業	自治体・都市計画提案事業者	• 利用者データ • 移動経路データ	自治体の都市計画における駐輪場、自転車・歩行者の通行帯の検討に利用

【出所】匿名加工情報の事例集［一般財団法人 日本情報経済社会推進協会］に基づき作成
https://www.jipdec.or.jp/protection_org/u71kba00000001hh-att/AOP_006.pdf

- ウェアラブルデバイスをはじめとするIoTや医療機関から得られた健康に関するデータを匿名加工情報とすることで、ヘルスケア分野における活用も期待されています。

匿名加工情報の作成・取り扱いに関する法令・ガイドライン

◆匿名加工情報の作成・利用に関する法令やガイドラインが示されています。

- 匿名加工情報は、**特定の個人を「識別」「復元」できないように、注意深く作成する必要**があります。
  - 匿名加工情報の利用面においても、データの作成元となった個人を識別しようとする行為は、個人情報保護法第38条で禁止されています。
- 表に示す匿名加工情報を作成するための手法例は、個人情報保護委員会のガイドラインに示されています。

匿名加工情報の加工に係る手法例

手法名	解説
項目削除／レコード削除／セル削除	加工対象となる個人情報データベース等に含まれる個人情報の記述等を削除するもの。例えば、年齢のデータを全ての個人情報から削除すること（項目削除）、特定の個人の情報を全て削除すること（レコード削除）、又は特定の個人の年齢のデータを削除すること（セル削除）。
一般化	加工対象となる情報に含まれる記述等について、上位概念若しくは数値に置き換えること又は数値を四捨五入などして丸めることとするもの。例えば、購買履歴のデータで「きゅうり」を「野菜」に置き換えること。
トップ（ボトム）コーディング	加工対象となる個人情報データベース等に含まれる数値に対して、特に大きい又は小さい数値をまとめることとするもの。例えば、年齢に関するデータで、80歳以上の数値データを「80歳以上」というデータにまとめること。
ミクロアグリゲーション	加工対象となる個人情報データベース等を構成する個人情報をグループ化した後、グループの代表的な記述等に置き換えることとするもの。
データ交換（スワップ）	加工対象となる個人情報データベース等を構成する個人情報相互に含まれる記述等を（確率的に）入れ替えることとするもの。
ノイズ（誤差）付加	一定の分布に従った乱数的な数値を付加することにより、他の任意の数値へと置き換えることとするもの。
疑似データ生成	人工的な合成データを作成し、これを加工対象となる個人情報データベース等に含ませることとするもの。

【出所】個人情報の保護に関する法律についてのガイドライン（匿名加工情報編）［個人情報保護委員会事務局］
https://www.ppc.go.jp/files/pdf/guidelines04.pdf

- **匿名加工情報自体**に加えて、**匿名加工情報への加工方法の情報にも、安全管理措置**が課されています。

匿名加工情報の提供時の公表義務

◆匿名加工情報を提供する場合は、「情報項目」や「提供方法」を公表する義務があります。

- 匿名加工情報の提供する場合は、**「個人に関わる情報項目」「提供方法」**を提供を行う事業者のウェブサイト等を通じて公表する義務があります。
  - 匿名加工情報の作成のみを行い、第三者に提供していない場合には公表義務はありませんが、提供準備があることを広報する目的等から匿名加工情報の作成を公表しているケースもあります。
  - 匿名加工情報を提供する場合でも、提供先の事業者名を公表する必要はありません。

匿名加工情報の提供に関する公表項目・公表例

提供主体例	提供データ例	個人に関わる情報項目例（公表項目）	提供方法例（公表項目）
スーパーマーケット	販売履歴データ（ID-POS）	顧客の性別、年齢層（5歳単位）、購入した商品、購入年月、購入時間帯（3時間単位）	アクセス制限を設定したインターネット上のサーバにアップロード
医療機関	健康診断の受診データ	患者の性別、生年、身長（5cm単位）、体重（5kg単位）、受診履歴、受診年月	パスワードを設定した電子ファイルが格納されたDVDを配達記録をつけて郵送

- 匿名加工情報の提供先には、**提供するデータが匿名加工情報である旨を明示**する必要があります。
  - 匿名加工情報取扱事業者間の契約によっては、提供を受けた事業者が再び匿名加工情報を提供するケースが考えられますが、匿名加工情報の提供を繰り返した場合でも、当該データが匿名加工情報であるという情報は共有されなくてはなりません。

- 匿名加工情報を提供を受けた事業者は、識別行為の禁止義務や匿名加工情報の安全管理措置が課されますが、**匿名加工情報の提供を受けたことの公表は不要**です。
  - 匿名加工情報の提供を受けた事業者が、当該事業者内のみ匿名加工情報を分析・活用する場合は提供を受けた事実の公表の義務はありません。

匿名加工情報の作成・提供に関する公表事例

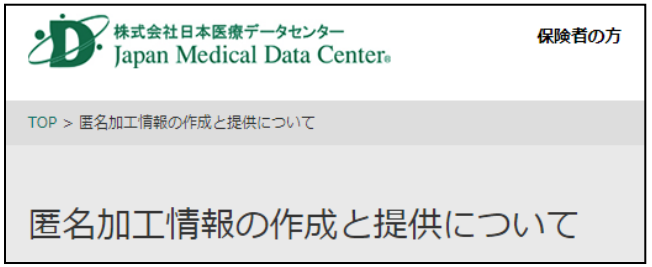
◆匿名加工情報の作成・提供情報は、インターネット等で公開されています。

- 匿名加工情報は「金融業」「小売業」「医療業」といった様々な業界で作成・第三者への提供が行われています。

匿名加工情報に関するインターネット上の公開例

【出所】株式会社イオン銀行
https://www.aeonbank.co.jp/privacy/rule/tokumei.html

【出所】コープデリ連合会
http://www.coopnet.jp/policy/tokumeikakou.html

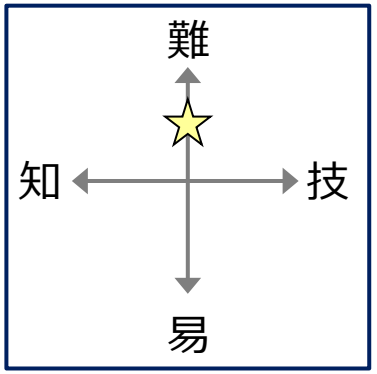


【出所】株式会社日本医療データセンター
https://www.jmdc.co.jp/other/anonymous.html

【出所】三井住友海上火災保険株式会社
http://www.ms-ins.com/privacy/anonymous.html

- 匿名加工情報制度によって匿名性を確保しながら、個人単位のデータを利活用することが可能になっています。

総務省　ICTスキル総合習得教材



［コース2］データ蓄積

2-5：多様化が進展するクラウドサービス

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_2_5.pdf

	1	2	3	4	5
［コース1］データ収集					
［コース2］データ蓄積					●
［コース3］データ分析					
［コース4］データ利活用					

本講座の学習内容［2-5：多様化が進展するクラウドサービス］

【講座概要】

- コンピュータの構成パーツと各パーツの基本的な役割を紹介します。
- クラウド上のGPUを活用する技術として、GPUクラウドコンピューティングを説明します。
- クラウド上の複数の物理サーバを一体として利用する分散技術を紹介します。
- crowd（群衆/大勢の人）に由来するクラウドサービスを紹介します。
- 国内企業におけるクラウドサービスの利用と労働生産性の関係を示します。

【講座構成】

座学

- [1] サーバの役割と構成パーツ
- [2] GPUクラウドコンピューティング
- [3] 分散技術
- [4] クラウドサービスの普及と効果

【学習のゴール】

- ✓ コンピュータを構成するパーツと各パーツの基本的な役割を把握する。
- ✓ GPUクラウドコンピューティングおよび分散技術の概要と用途を理解する。
- ✓ crowd（群衆/大勢の人）に由来するクラウドサービスを説明できる

多様なクラウドの活用

◆クラウドはデータ蓄積のみならず、様々な用途で活用されています。

- クラウドプラットフォームのサービスカテゴリにおいて「データ蓄積」に関するサービスは、全体の一部となっています。
  - 各サービスカテゴリはいくつかのクラウドサービスによって構成されており、クラウドサービスの数はカテゴリの数よりも多くなっています。

3大クラウドプラットフォームのサービスカテゴリ［2018年3月時点］

AWS（Amazon）	Azure（Microsoft）	GCP（Google）
コンピューティング	AI + 機械学習	コンピューティング
ストレージ	DevOps	ビッグデータ
データベース	ID	Identity & Security
移行	Storage	モノのインターネット（IoT）
ネットワークとコンテンツ配信	Web	専門サービス
開発者用ツール	コンテナー	ストレージとデータベース
管理ツール	コンピューティング	API プラットフォームとエコシステム
メディアサービス	セキュリティ	管理ツール
セキュリティ、アイデンティティ、コンプライアンス	データベース	ネットワーキング
分析	ネットワーク	Cloud AI
機械学習	メディア	データ移行
モバイルサービス	モノのインターネット (IoT)	デベロッパーツール
ARおよびバーチャルリアリティ	モバイル	生産性向上ツール
アプリケーション統合	移行	
カスタマーエンゲージメント	開発者ツール	
ビジネスの生産性	管理ツール	
デスクトップとアプリケーションのストリーミング	統合	
IoT	分析	
ゲーム開発		

「データ収集」と関連の深いカテゴリ

「データ蓄積」と関連の深いカテゴリ

「データ分析・処理」と関連の深いカテゴリ

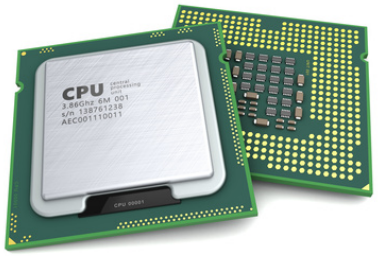
□ この講座では、クラウドにおける「データ分析・処理」に関する技術を中心に、多様なクラウドの活用を示します。

「データ分析・処理」に関するクラウドの技術

◆この講座では、「GPUクラウドコンピューティング」と「分散技術」を紹介します。

- 「データ分析・処理」に関するクラウドの技術として**GPUクラウドコンピューティング**と**分散技術**を紹介します。
  - クラウドを含むサーバコンピュータの従来の主目的は「長期的なデータ蓄積」でしたが、「データ分析、処理」を主目的とする使い方が現れています。
    - GPUクラウドコンピューティング（GPUを利用して高速のデータ処理を行う）
    - 分散技術（複数の物理サーバを一体として利用して、データの蓄積・処理を行う）
- **GPUクラウドコンピューティング**は、クラウド上の**GPU（画像処理装置）を利用して高速のデータ処理**を行います。
  - 通常、コンピュータの処理において「データ分析・処理」を担うのは、「CPU（Central Processing Unit：中央処理装置）」ですが、GPUコンピューティングでは「データ分析・処理」に「GPU（Graphics Processing Unit：画像処理装置）」を利用します。
  - コンピュータの中には、CPUとGPUの両方が入っており、起動時には両方が稼働しています。

CPU（中央処理装置）



GPU（画像処理装置）
が搭載されたグラフィックボード

- **分散技術**は、**複数の物理サーバを一体として利用して大容量のデータの蓄積・処理**を行う技術です。
  - 仮想化技術は、物理サーバの中に仮想サーバを構築する使い方が一般的ですが、分散技術は物理サーバの枠を広げる技術です。

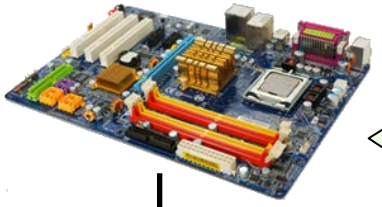
□ まず、コンピュータパーツの構成から「GPUクラウドコンピューティング」を示し、続いて「分散技術」を紹介します。

コンピュータを構成するパーツ

◆物理サーバ（サーバコンピュータ）やPCは、同様のパーツで構成されています。

- 物理サーバ（サーバコンピュータ）でも、家庭等で利用されるPCでも、コンピュータを構成する主要パーツは共通しており、**マザーボード**、**CPU**、**メインメモリ**、**HDD**、**グラフィックボード**の5種で構成されています。

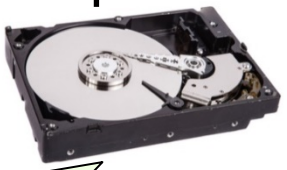
PCのパーツ構成と一般的な利用における各パーツの役割



マザーボード
コンピュータパーツを連結し、
データのやりとりを仲介

CPU
（中央処理装置）
データ分析・処理

メインメモリ
一時的なデータ蓄積
［CPUとのデータのやりとりは高速］



HDD（ハードディスクドライブ）
長期的なデータ蓄積
［CPUとのデータのやりとりは低速］

グラフィックボード
グラフィックボード上のGPUで
画像処理・ディスプレイへの表示

※ 各パーツの写真は異なるサイズ調整をしているため、実際のパーツにおける大きさの関係は、これらの写真における大きさの関係とは異なります。

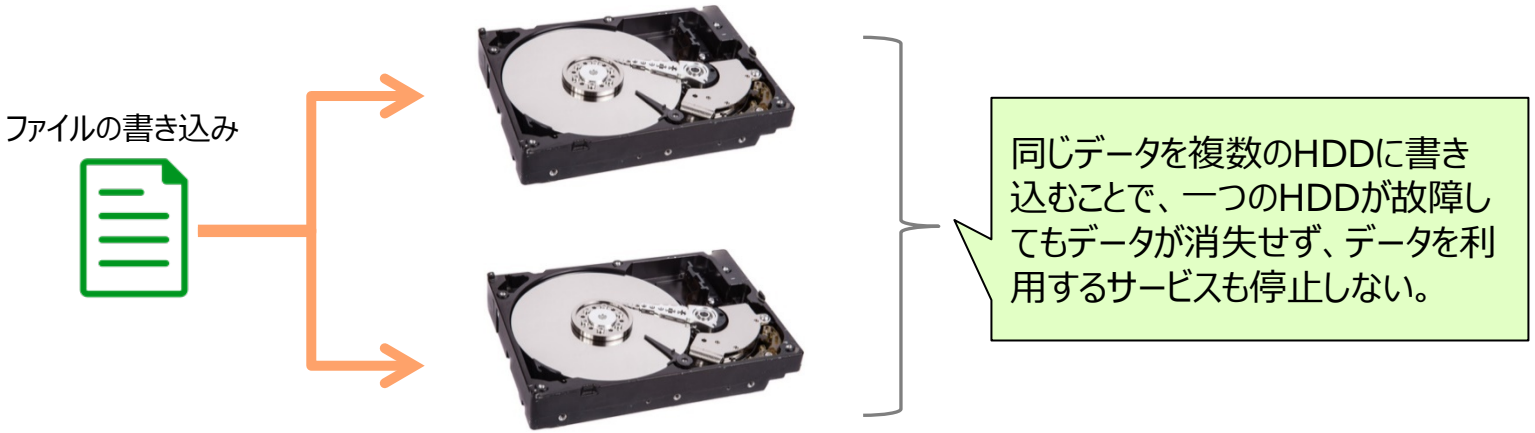
- メインメモリに蓄積したデータは電源を切ると消えますが、ハードディスクドライブに蓄積したデータは電源を切っても消えません。
- 長期的なデータ蓄積として、磁気ディスクに記録するハードディスクドライブの代わりに、半導体メモリに記録するSSD(ソリッドステートドライブ)を利用するケースもあります。
- グラフィックボードは個別のパーツになっておらず、マザーボード上に一体化している「オンボードグラフィック」のケースもあります。
- 枠内のパーツ写真はデスクトップPC用のパーツですが、ノートPCではこれらのパーツが小型化、凝縮される形で機材に格納されています。

サーバコンピュータの基本的かつ重要な役割

◆サーバコンピュータの基本的かつ重要な役割は、HDDが担う「長期的なデータ蓄積」です。

- サーバコンピュータ上のデータベースには企業の「顧客情報」「決済情報」といった消失や誤りが許されないデータを蓄積しているケースもあります。
  - 講座2-1に示したように、サーバコンピュータ上のデータベースには大容量のデータが蓄積され、容易に検索できることで、その価値を発揮します。
- サーバコンピュータでは、一般のPCに比べても、安全かつ長期的なデータ蓄積に配慮しています。
  - 前スライドに示したように「長期的なデータ蓄積」を担うコンピューターパーツは、HDDです。
- データセンター内のサーバコンピュータは、一般に複数のHDDが格納されており、一部のHDDが故障してもデータが消失しないよう多重化しています。
  - 1台のコンピュータ内にある複数のHDDに同時に読み書きする技術をRAID(Redundant Arrays of Independent Disks：レイド)と言います。
  - RAIDには複数のタイプがありますが、ミラーリングとも呼ばれるRAID1では、同じデータを複数のHDDに書き込みます。

RAID1（ミラーリング）を持ったHDDの利用



□ 多様なクラウドサービスが普及しつつありますが、サーバコンピュータとして「長期的なデータ蓄積」は継続して重要です。

2種類のHPC（ハイパフォーマンスコンピューティング）

◆HPC（ハイパフォーマンスコンピューティング）にはCPU活用型とCPU活用型があります。

- サーバコンピュータの最も重要な役割は**「長期的なデータ蓄積」**であることが一般的でした。
  - ウェブサーバ、メールサーバ、ファイルサーバ、データベースサーバといった用途では、データの処理速度よりも安全かつ「長期的なデータ蓄積」が重要でした。
  - サーバコンピュータ内のパーツとしても「長期的なデータ蓄積」を担うHDDが特に重要視され、HDDの情報セキュリティに着目されていました。
- 近年、サーバコンピュータを**「データ分析・処理」**のために利用する**HPC（ハイパフォーマンスコンピューティング）**が着目を浴びています。
  - HPCは、スーパーコンピュータで実施するような高速計算やPCやスマートフォンで時間がかかる計算をサーバ上で処理する利用が考えられます。
- HPCには、**CPUを利用するケース**とグラフィックボード上の**GPUを利用する「GPUコンピューティング」**のケースの2種類があります。
  - GPUの従来の目的は画像処理と表示でしたが、GPUコンピューティングでは画像データに限らず、あらゆるデータをGPUにて処理、計算します。

サーバ用途：長期的なデータ蓄積

◆ ウェブサイトやデータベースの運用

特に重要なPCパーツは
HDD（ハードディスクドライブ）

サーバ用途：データ分析・処理

◆ HPC（ハイパフォーマンスコンピューティング）

CPU活用のケース

特に重要なPCパーツは
CPU（中央処理装置）

GPU活用のケース（GPUコンピューティング）

特に重要なPCパーツは
グラフィックカード

スーパーコンピュータ（CPU活用型・GPU活用型）

◆HPCの代表例として、スーパーコンピュータが挙げられます。

- 理化学研究所と富士通株式会社が共同開発したスーパーコンピュータ「京（けい）」は、2017年11月の国際的なHPC（ハイパフォーマンスコンピューティング）ランキングにおいて、3期連続で世界第1位を獲得しました。

  【出所】「京」が性能指標(HPCG)において3期連続で世界第1位を獲得（理化学研究所）
  http://www.riken.jp/pr/topics/2017/20171115_1/

- 東京工業大学のスーパーコンピュータ「TSUBAME（つばめ）3.0」は、2017年6月における省エネ性能の世界スパコンランキングであるGreen500 Listにおいて、世界第1位を獲得しました。

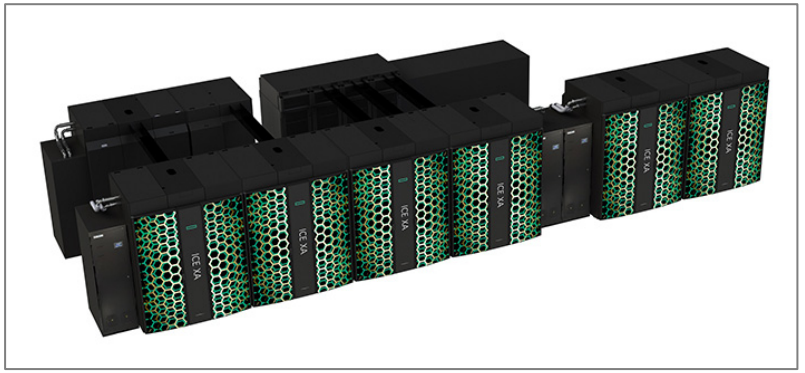
  【出所】東工大TSUBAME3.0と産総研AAICが省エネ性能スパコンランキングで世界1位・3位を獲得（東京工業大学）
  https://www.titech.ac.jp/news/2017/038640.html

- 「京」は、8万8128個の高性能CPUによって構成されているCPU活用型のスーパーコンピュータである一方で、「TSUBAME 3.0」は、2160枚のGPUを利用するGPU活用型のスーパーコンピュータです。
  - TSUBAMEのGPUは、1個あたり3584個のGPUコアがあるため、合計774万1440個のGPUコアを計算に利用可能です。

京（CPU活用型）

【出所】理化学研究所 計算科学研究センター
http://www.r-ccs.riken.jp/jp/k/

TSUBAME3.0（GPU活用型）



【出所】東京工業大学
https://www.titech.ac.jp/news/2017/037500.html

□ こうしたスーパーコンピュータも、講座2-1に示したNISTの5つの要件を満たせば、クラウドと考えることができます。

CPUと比較したGPUの特徴

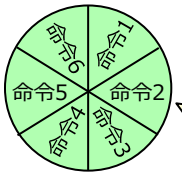
◆GPUはCPUに比べて、クロック周波数は小さい一方で、コアの個数が多くなっています。

- コンピュータの計算速度はCPUやGPUの中にある**コア（計算ユニット）のクロック周波数**と**コアの個数**に依存します。
- **クロック周波数**が大きければ、**単位時間あたりに実行できる命令数**が増加し、計算速度が高まります。
  - クロック周波数が4.0GHzという表記は、1秒間に命令を行う周期に当たるクロックが40億回あることを意味しています。

クロック周波数が小

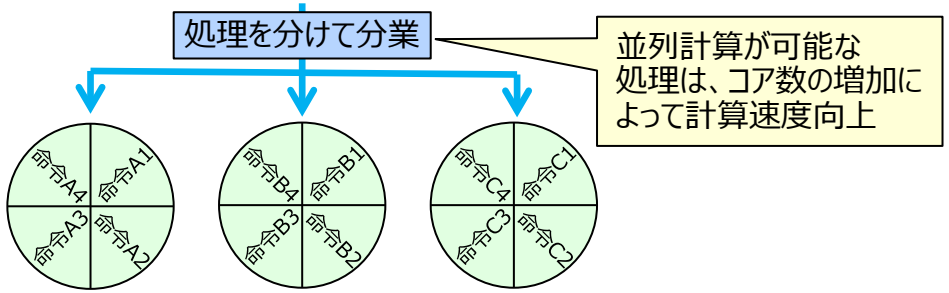
単位時間で実行可能な命令数が少なく、計算速度が遅い

クロック周波数が大



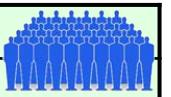
単位時間で実行可能な命令数が多く、計算速度が速い

- **コアの個数**が多ければ、複数のコアで**分業する並列計算が可能な処理**に関しては、計算速度が高まります。
  - Intel社のCPUでは、コア内で処理（スレッド）をさらに分けるハイパースレッディング・テクノロジーが搭載されているケースもあります。



- **GPUは**CPUに比べて、クロック周波数は小さい一方で**コアの個数が多い傾向にあり、並列計算に適しています。**

一般PC向けのCPUとGPUの比較

	CPU Core i7-6700K	GPU GEFORCE GTX 1080 
コアのクロック周波数（ターボ時）	4.0GHz（4.2GHz）	1.607GHz（1.733GHz）
コアの個数（スレッド数）	4(8)	2560
製造者（製品の発売時期）	Intel（2015年8月）	NVIDIA（2016年5月）

【出所】Intel: https://ark.intel.com/ja/products/88195/Intel-Core-i7-6700K-Processor-8M-Cache-up-to-4_20-GHz
NVIDIA: https://www.nvidia.com/en-us/geforce/products/10series/geforce-gtx-1080/

- CPUは少数精鋭部隊のイメージであるのに対して、GPUは軍団のイメージとなります。

GPUクラウドコンピューティング

◆「GPUクラウドコンピューティング」は、クラウド上のGPUを活用するHPCです。

- CPUは個々のコアのクロック数が周波数が高いため、順番に行う処理に適している一方で、GPUはコアの個数が多いため、分業（並列計算）ができる処理に適しています。
- 人口知能の分析技術である機械学習は、並列計算によって大幅に高速化できるケースがあり、有力なGPU開発・製造企業であるNVIDIA（エヌビディア）社は、GPUが機械学習に適していることを論じています。
  - ディープラーニング等の機械学習の手法や活用のイメージは、講座3-5において示します。
- クラウド上の高性能GPUを利用してHPC（ハイパフォーマンスコンピューテング）を行うことを**「GPUクラウドコンピューティング」**といいます。

NVIDIA社による機械学習（GPUアプリケーション）の紹介

【出所】機械学習GPU で加速化したアプリケーション［NVIDIA］
http://www.nvidia.co.jp/object/machine-learning-jp.html

NVIDIA社によるGPUクラウドコンピューティング紹介

【出所】GPUクラウドコンピューティング［NVIDIA］
http://www.nvidia.co.jp/object/gpu-cloud-computing-jp.html

GPUクラウドコンピューティングのサービス例

◆3大クラウドプラットフォームでは、それぞれGPUクラウドコンピューティングを提供しています。

- AWS、Azure、GCPは、GPUクラウドコンピューティングに関するサービスを2016年度に増強・新設しました。

3大クラウドプラットフォームが提供するGPUクラウドコンピューティングに関するサービス

AWSのGPU活用

【出所】AWS［Amazon］
https://aws.amazon.com/jp/blogs/news/new-p2-instance-type-for-amazon-ec2-up-to-16-gpus/

AzureのGPU活用

【出所】Azure［Microsoft］
https://blogs.technet.microsoft.com/jpitpro/2016/11/22/azure-n-series-general-availability-on-december-1/

GCPのGPU活用

【出所】GCP［Google］
https://cloudplatform-jp.googleblog.com/2016/11/gpu-2017.html

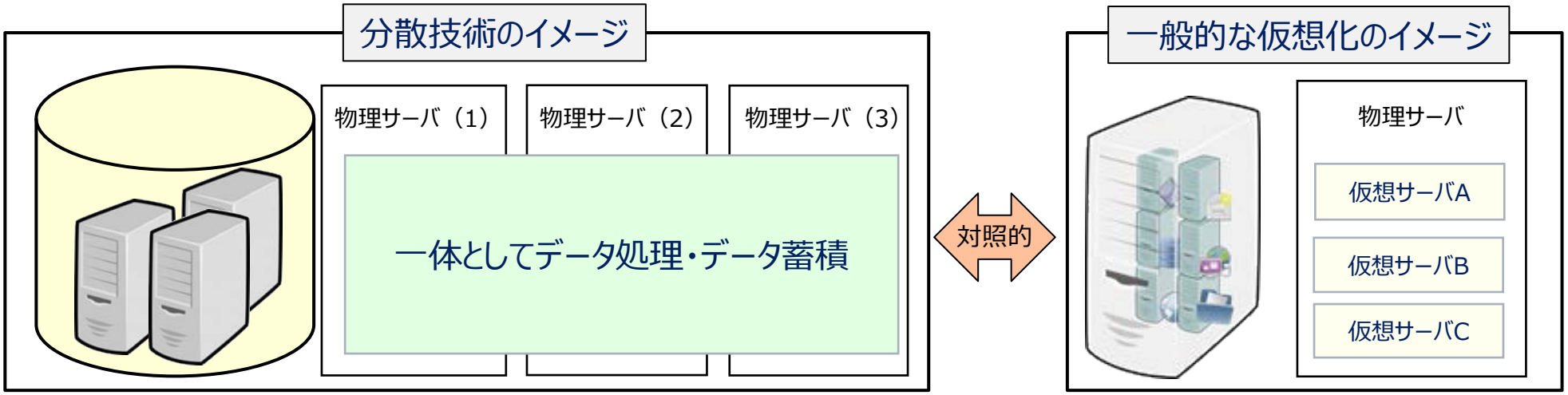
- AWS、Azure、GCPが提供するGPUクラウドコンピューティングのサービスは、1時間単位で利用することができます。
  - 高性能GPUを利用する場合であっても、短時間であれば安価に利用することができます。

□ 以上の「GPUクラウドコンピューティング」の説明に続いて、「分散技術」の説明に移ります。

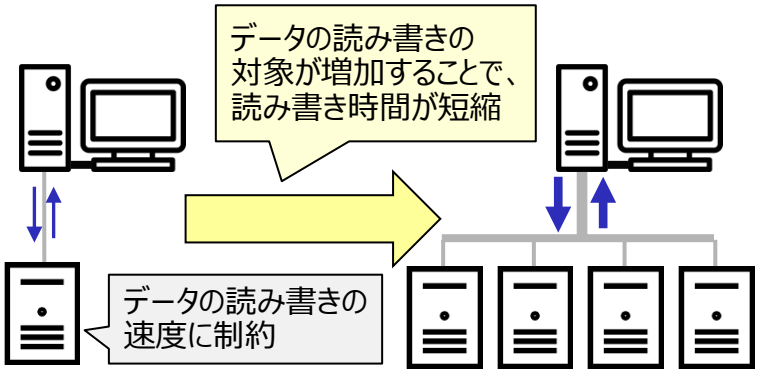
分散技術のイメージ

◆分散技術は、複数のコンピュータを一体として利用する形で、物理的なコンピュータの枠にとらわれない技術です。

- **分散技術**は、一般的な仮想化とは対照的に**複数のコンピュータを一体として利用**することで、データ蓄積やデータ処理を分散する手法です。
  - 講座2-1で紹介したように、一般的な仮想化では、1台の物理サーバの中に複数の仮想サーバを構築します。
  - 分散技術によって複数のコンピュータや機器を一体として利用することで、個々のハードウェアの機能を超えた性能や容量を利用することができます。



- 分散技術を用いることで、GB（ギガバイト）、TB（テラバイト）、PB（ピコバイト）単位の大容量データの蓄積、処理が可能となります。
- コンピュータやHDDの個数が増え、データの読み書きの対象が増加することで、読み書きの時間が短縮される効果もあります。
  - 本講座で紹介した「1台のコンピュータ内にある複数のHDDに同時に読み書きする技術」のRAIDには、RAID0（ストライピング）など、読み書き速度を速めるタイプもあります。



代表的な分散技術：Hadoop

◆分散技術として、複数のサーバにデータと処理を分散させるHadoopが代表的です。

- 分散技術はApache（アパッチ）ソフトウェア財団が開発し、無償利用できる**Hadoop**（ハドゥープ）が代表的です。
  - Apacheソフトウェア財団は、ウェブサーバ運営ソフトウェアをはじめとして無償利用が可能なソフトウェアを開発・提供している非営利団体です。
- Hadoopは、「国立国会図書館サーチのシステム」にも活用されています。

> 国立国会図書館サーチでは、データ変換、書誌同定処理、グループ化処理、インデックス作成に、Hadoopを使用しています。Hadoopの分散処理機構により、高速・効率的に処理を行なっています。

【出所】国立国会図書館サーチのシステムについて（国立国会図書館） http://iss.ndl.go.jp/information/system/

- Hadoopにおいてサーバは、データ処理を指揮するマスターサーバと処理作業を行うスレーブサーバ群に分かれます。
  - マスター(master)には、「主人、管理者、指示を出す者」、スレーブ（slave）には、「奴隷、管理される者、指示を受ける者」という意味があります。
- Hadoopの処理プロセスは、第1段階の分割保存を行う**HDFS（Hadoop Distributed File System）**と第2段階の分散処理技術の**MapReduce**と大きく2段階に分かれます。
  - 後述するようにMapReduceは、さらにMap処理とReduce処理に分けられ、その間にはShuffle処理が自動的に行われます。

Apache Hadoopのロゴ

Copyright © 2017 The Apache Software Foundation

【出所】Apache Hadoop［The Apache Software Foundation］
http://hadoop.apache.org/

2種類のサーバ

- マスターサーバ
- スレーブサーバ（群）

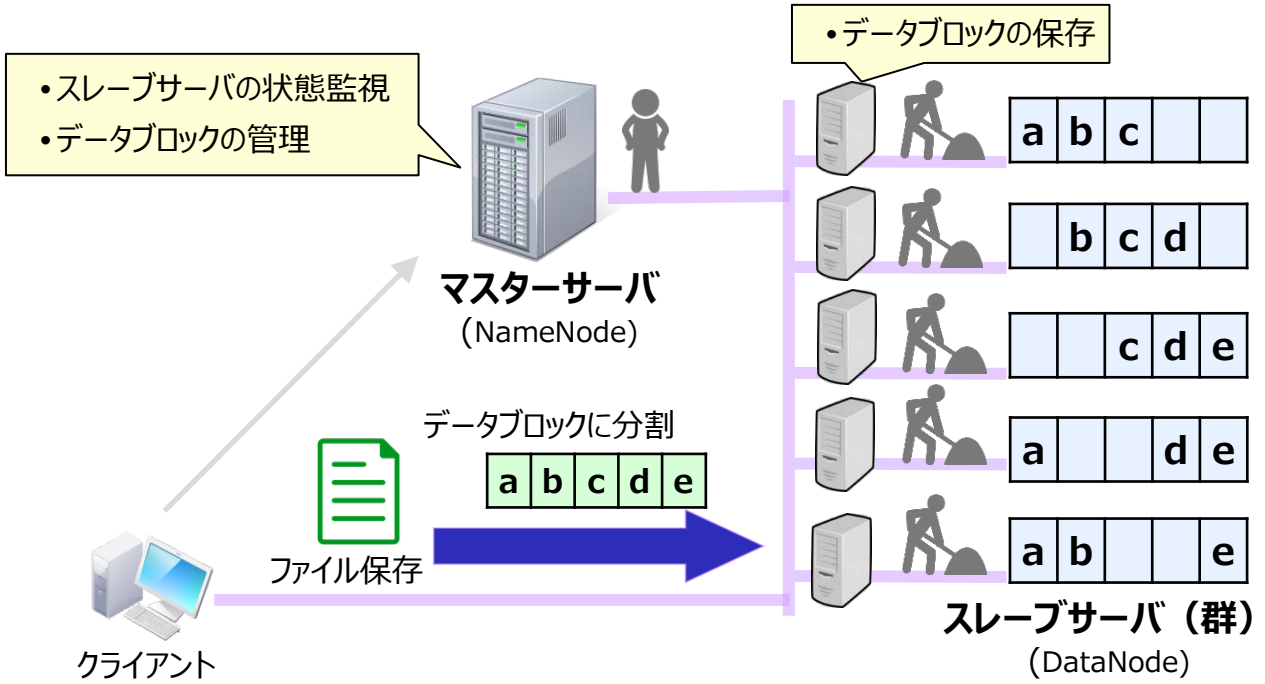
2段階の処理プロセス

- HDFS
- MapReduce
  - Map処理
  - （Shuffle処理）
  - Reduce処理

Hadoopの第1段階：HDFS（分割保存）

◆HDFSは、Hadoopの1段階目に当たるデータの分割保存です。

- **HDFS**（Hadoop Distributed File System）は、複数のコンピュータにデータを**分割保存する枠組み**です。
  - 同一コンピューター内で複数のHDDに保存する技術（RAID）とは異なり、HDFSはコンピュータをまたいでデータを分割保存します。
- マスターサーバは、スレーブサーバ群にある個々のデータブロックの位置・名前を管理し、読み書きの指示を与えます。
  - HDFSでは、ファイルを一定のデータサイズ（初期設定では64MB）に分割したブロックにして、スレーブサーバ群での読み書きを行います。
  - マスターサーバはデータブロックの名前を管理することからNameNode、スレーブサーバはデータブロックを取り扱うことからDataNodeとも呼ばれます。
- HDFSはスレーブサーバの故障にも配慮しており、スレーブサーバ群のデータブロックは定められた数の複製を作ります。
  - 平均5年（60ヶ月）に1度故障するスレーブサーバを60台連結して、データブロックの複製せずに分割保存すれば、確率的には毎月平均1台のスレーブサーバの故障が発生し、データ全体は平均1ヶ月で使用不可能になります。



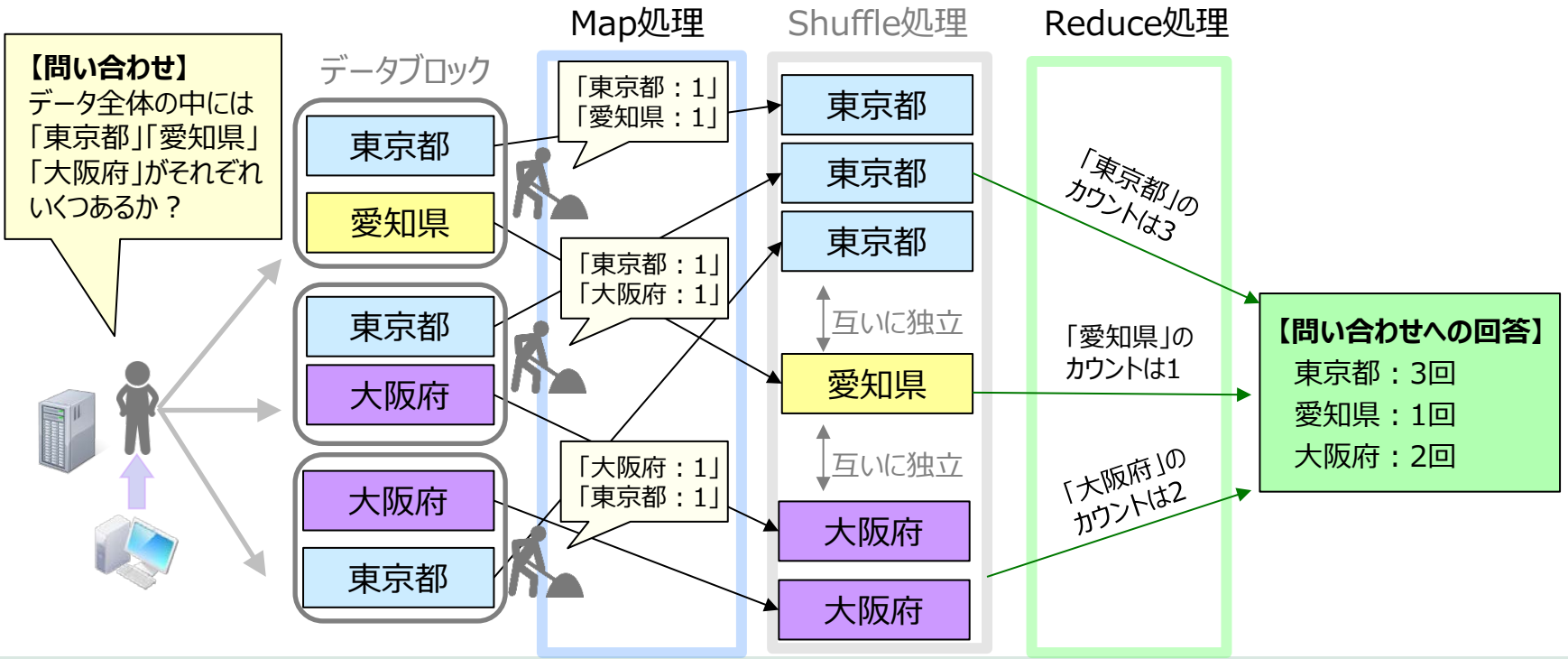
◆ 左図において、一部のスレーブサーバが故障しても、データブロック a b c d e の全てを利用できれば、データ全体を復元することができます。

◆ データ全体の一部を個々のスレーブサーバに保存することで、データ全体を一つのサーバに保存する場合に比べて、読み書きにかかる時間を短縮することができます。

Hadoopの第2段階：MapReduce（分散処理技術）

◆MapReduceは、Hadoopの2段階目などで利用される分散処理技術です。

- **MapReduce**は、Map処理とReduce処理の順に行われる**分散処理技術**です。
  - MapReduceは分散処理技術の一般的な名称であり、Hadoopの2段階目の処理以外でも活用されています。MapReduceは、2004年にGoogleに所属する研究者から論文が発表されました。
    【出所】MapReduce: Simplified Data Processing on Large Clusters
    ［Jeffrey Dean and Sanjay Ghemawat｜Google］https://ai.google/research/pubs/pub62
- Map処理では、各スレーブサーバに分散して保存されたデータブロックを読み込み、必要なデータを抽出・変換します。
  - Hadoopでは概ね自動化されていますが、Map処理とReduce処理の間には、抽出したデータを並べ替えてまとめるShuffle処理が入ります。
- Reduce処理では、抽出・変換されたデータに対して目的の作業を行い、出力します。



新しい分散処理技術：Apache Spark

◆Apache Sparkをはじめとして、新しい分散処理技術が注目を集めています。

- **Apache Spark（スパーク）**は、Hadoopの第2段階のMapReduceにおける短所を克服する形で誕生しました。
  - Sparkにおいても、Hadoopの第1段階のHDFSを活用することは、従来のHadoopと共通しています。
- Sparkの利点として、**データへのアクセス速度**、**利用可能なプログラミング言語**、**データベースの利用**が挙げられます。

➢ Sparkは過去に利用したデータは、スレーブサーバのHDDではなく、メインメモリから読み出すことで**高速化**しました。
- 本講座の「コンピュータを構成するパーツ」で紹介したように、メインメモリはHDDに比べてCPUとのデータのやりとりを高速に行うことができます。
- Apacheソフトウェア財団では、データ処理の構造を変えることでHDD上にあるデータに関しても、Sparkの方がアクセス速度が速いとしています。

➢ Sparkでは、Java, Scala, **Python**, **R**と複数のプログラム言語を利用して分散処理を実行できます。
- Pythonはディープラーニング等の機械学習を実行する言語としても人気があり、講座4-3に示すRはデータ分析用の言語として人気があります。

➢ Sparkでは、**データベースの利用に関する利便性も向上**しました。
- Hadoopでのデータベース利用はデータベースソフトウェアを追加インストールする必要がありますが、SparkにはSpark SQLが標準で付属しています。

Apache Sparkのロゴ

Copyright © 2017 The Apache Software Foundation

（Apache Hadoopの）MapReduceとSparkの比較

	（Apache Hadoopの）MapReduce	Apache Spark
データへのアクセス速度	Sparkと比べて低速（スレーブサーバのHDDに毎回アクセスする必要があるため）	利用済データへのアクセス速度は特に速い（利用したデータは、容量上限までメインメモリに保存されるため）
利用可能なプログラム言語	Java	Java, Scala, Python, R
データベースの利用	Hive、Hbaseの追加導入の必要性	Spark SQLが標準で付属

【出所】Apache Spark［The Apache Software Foundation］の資料に基づいて作成 https://spark.apache.org/

クラウドにおける分散技術のサービス例

◆3大クラウドプラットフォームでは、分散技術を活用したクラウドサービスを提供しています。

- AWS、Azure、GCPでは、それぞれ分散技術を活用したクラウドサービスを提供しています。

3大クラウドプラットフォームが提供する分散技術を活用したクラウドサービス

Amazon EMR

【出所】Amazon EMR［AWS｜Amazon］
https://aws.amazon.com/jp/emr/

HDInsight

【出所】HDInsight ［Azure｜Microsoft］
https://azure.microsoft.com/ja-jp/services/hdinsight/

CLOUD DATAPROC

【出所】CLOUD DATAPROC［GCP｜Google］
https://cloud.google.com/dataproc/

- 分散技術はHDFSによって安全なデータ蓄積を可能にすることに加えて、分散処理技術によってプログラミングやデータベース利用における高速のデータ処理を可能としています。

□ 以上で「データ分析・処理」に関わるクラウド利用の説明を終了し、より広い観点からのクラウド利用を紹介します。

クラウドとIoTの親和性

◆3大クラウドプラットフォームでは、IoT活用を目的としたクラウドサービスを提供しています。

- クラウドはIoTとの親和性が高く、IoT活用を目的としたクラウドサービスも提供されています。
- AWS、Azure、GCPでは、それぞれIoT活用を目的とするクラウドサービスを提供しています。
  - 本講座の冒頭に示した「3大クラウドプラットフォームのサービスカテゴリ」の表でも、IoT活用を目的としたカテゴリはセルを青色していました。

3大クラウドプラットフォームが提供するIoT活用を目的とするクラウドサービス

AWS IoT Core

【出所】AWS IoT Core［AWS｜Amazon］
https://aws.amazon.com/jp/iot-core/

Azure IoT Hub

【出所】Azure IoT Hub［Azure｜Microsoft］
https://azure.microsoft.com/ja-jp/services/iot-hub/

Google Cloud IoT

【出所】Google Cloud IoT ［GCP｜Google］
https://cloud.google.com/solutions/iot/

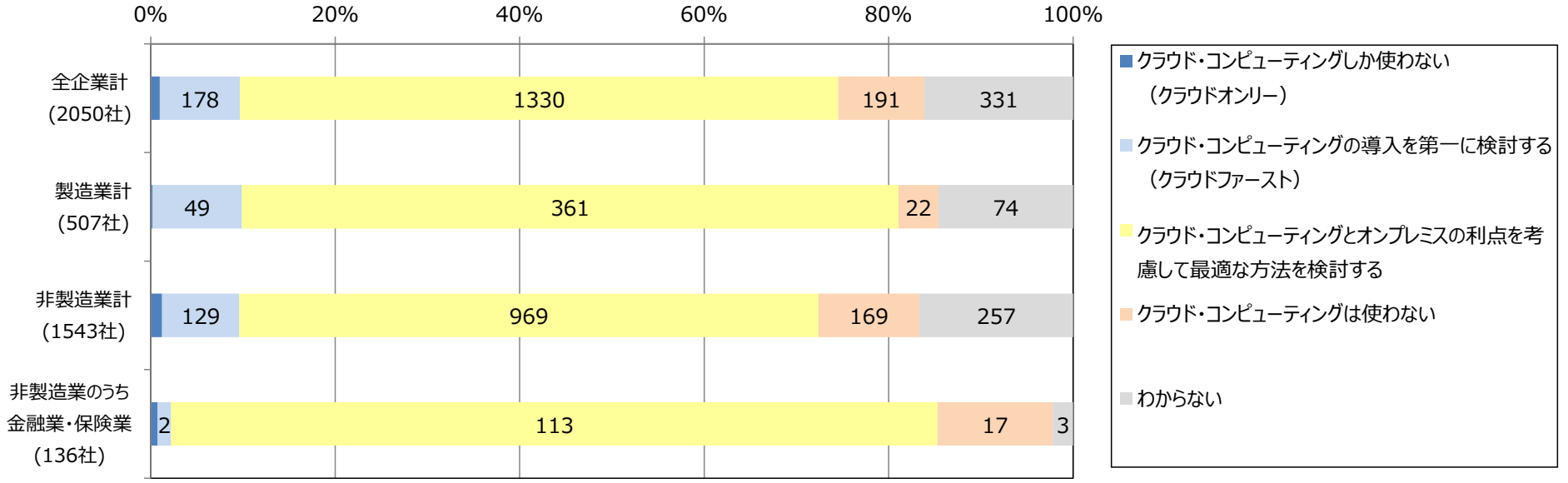
□ 本講座では、コース3と関わりが深い「データ分析・処理」に関わるクラウド利用を紹介してきましたが、コース1と関わりが深い「データ収集・IoT活用」としても、クラウドは幅広く活用することができます。

クラウドの導入を第一に検討（クラウドファースト）

◆クラウドの導入を第一に検討する「クラウドファースト」の意識を持つ企業もあります。

- 経済産業省が公表する「情報処理実態調査」では、民間企業のクラウドの利用に関する意識等を調査しています。
- 情報処理実態調査では［クラウド・コンピューティングしか使わない］ことを**「クラウドオンリー」**とし、［クラウド・コンピューティングの導入を第一に検討する］ことを**「クラウドファースト」**として、回答項目を設定しています。

「貴社では、今後の情報システムのへの取り組みに関して、クラウド・コンピューティングの導入をどのように位置づけていますか。」の回答割合



【出所】平成28年 情報処理実態調査［経済産業省］の資料に基づき作成 http://www.meti.go.jp/statistics/zyo/zyouhou/index.html

- 製造業では「クラウドファースト」の回答割合は「クラウド・コンピューティングは使わない」を上回っています。
- 非製造業では「クラウドファースト」の回答割合は「クラウド・コンピューティングは使わない」を下回っています。
  - 非製造業の中でも、金融・保険業では「クラウド・コンピューティングは使わない」と回答する企業の割合が特に高くなっています。

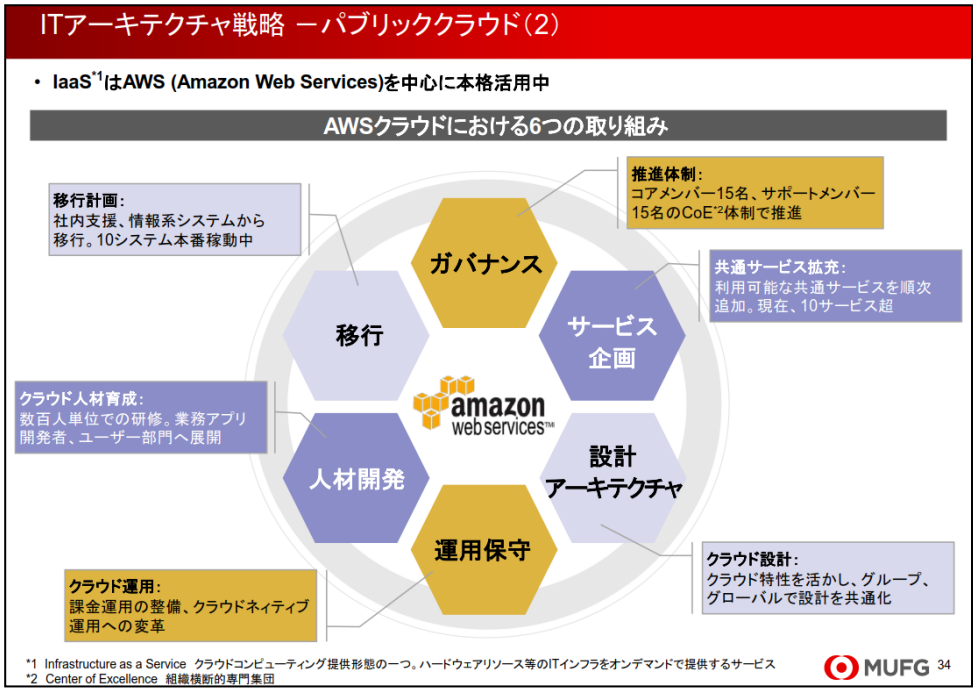
□ 金融・保険業を中心に、非製造業ではクラウドの利用に距離を置く傾向にありましたが、状況は変わりつつあります。

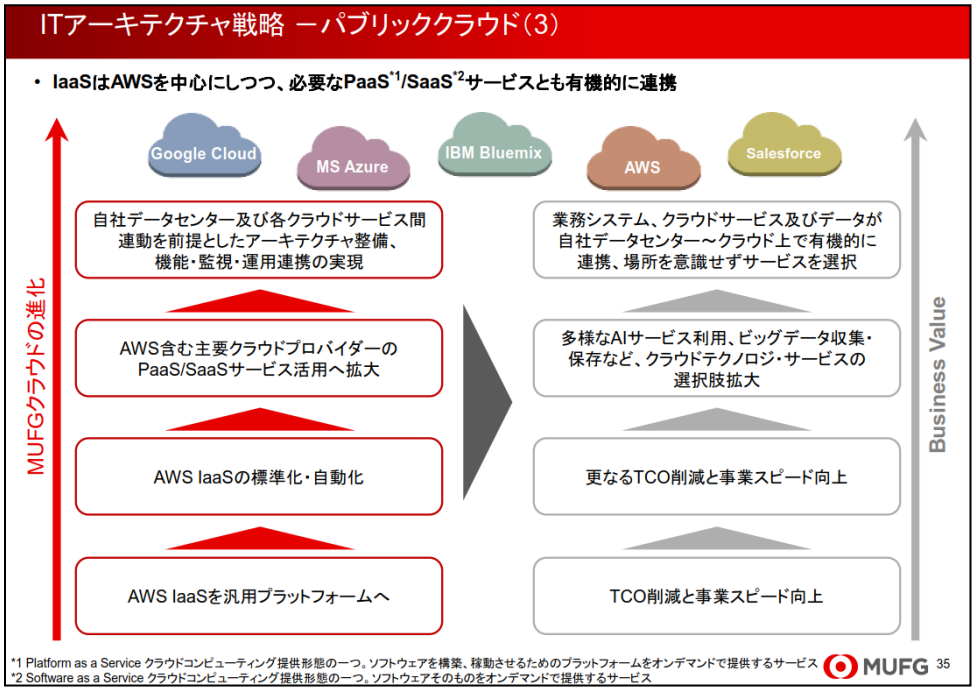
クラウドを活用する企業事例（金融・保険業）

◆金融・保険業においても、クラウドの利用が進展しつつあります。

- 株式会社三菱UFJフィナンシャル・グループおよびそのグループ会社では、クラウドの利用を推進しています。
- クラウドの各サービスモデルや各事業者のサービスの長所を活かす形でのクラウド利用が進展しています。
  - 各サービスモデル（IaaS、PaaS、SaaS）に関しては、講座2-2にて説明しました。
  - クラウドサービスの中には講座1-5で説明したAPIに関連するサービスもあります。また、講座4-4では「銀行のオープンAPIでの情報連携」を紹介します。
- 他の業界に比べても保守的とされていた金融・保険業においても、クラウドの利用は進展しています。

株式会社三菱UFJフィナンシャル・グループのデジタルトランスフォーメーション戦略の資料（抜粋）

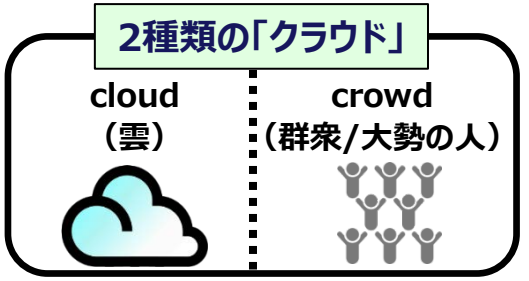




【出所】デジタルトランスフォーメーション戦略（2017年9月11日）［株式会社三菱UFJフィナンシャル・グループ］
https://www.mufg.jp/ir/presentation/backnumber/pdf/slides170911.pdf

crowd（群衆/大勢の人）としてのクラウド

◆crowd（群衆/大勢の人）に由来するクラウドサービスもあります。



- 本コースで説明してきたクラウドは、**「雲」を意味するcloud**です。
  - 「クラウド コンピューティング」の英語表記は［cloud computing］です。
- クラウドには**「群衆/大勢の人」を意味するcrowd**が使われるケースもあります。
- crowdという言葉が利用されるサービスとして、大勢の人に知恵や支援を求める**クラウドソーシング**、大勢の人に資金援助を求める**クラウドファンディング**が挙げられます。
  - 「クラウドソーシング」や「クラウドファンディング」の英語表記は、それぞれ［crowdsourcing］と［crowdfunding］です。

クラウドソーシングサービスの「CrowdWorks （クラウドワークス）」

【出所】CrowdWorks（クラウドワークス）https://crowdworks.jp

クラウドファンディングサービスの「Makuake （マクアケ）」

【出所】Makuake（マクアケ）https://www.makuake.com/

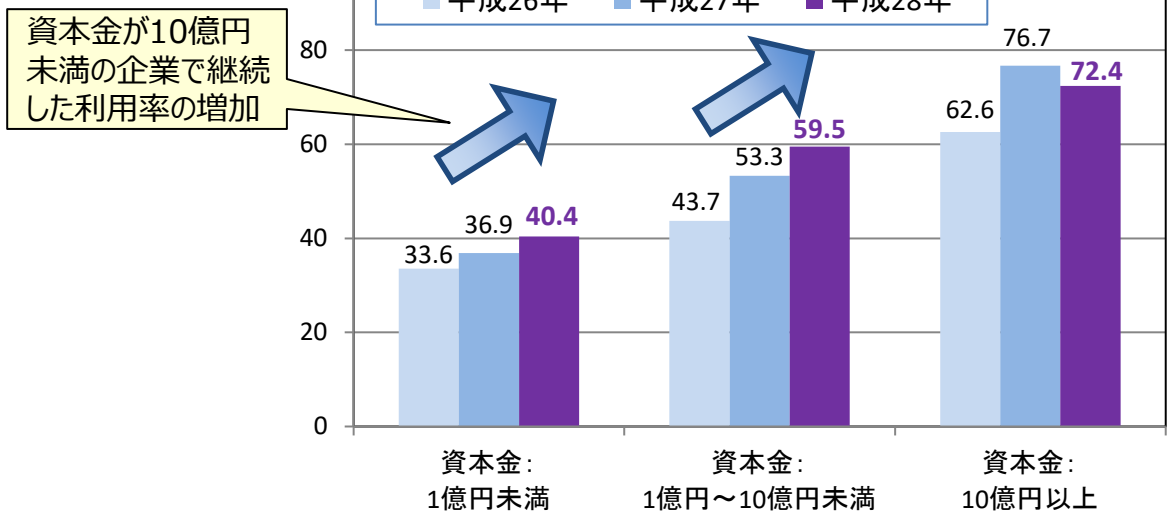
□ cloud（雲）サービスの普及によって、crowd（群衆/大勢の人）に関するサービスにも使われるようになりました。

クラウドサービスの利用と企業の生産性

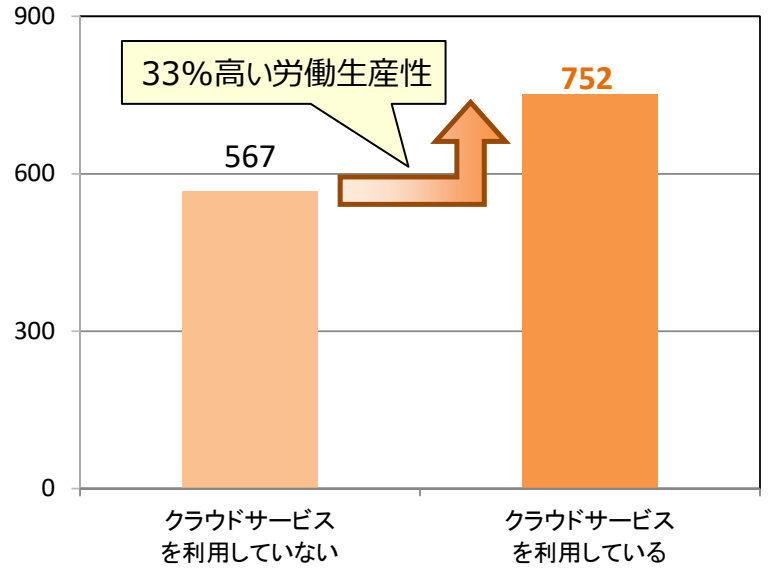
◆クラウドサービスの利用企業は「生産性が高い」という調査結果が示されています。

- 総務省が公表している通信利用動向調査（企業調査）の調査結果資料では、「資本金規模別のクラウドサービスの利用状況」や「クラウドの利用と生産性の関係」を示しています。
- 「資本金規模別のクラウドサービスの利用状況の推移」を見ると、資本金規模が大きい企業の方がクラウドサービスの利用割合の水準は高くなっていますが、資本金規模が小さい企業のクラウド利用率が継続して増加しています。
- 平成28年の通信利用動向調査（企業調査）の結果資料では、**クラウドサービスの利用がある企業は利用していない企業に比べて、労働生産性が33%高い**という結果を示しています。
  - この資料において、「労働生産性＝（営業利益＋人件費＋減価償却費）÷従業者数」として、導出しています。

（%） 資本金規模別のクラウドサービスの利用状況の推移



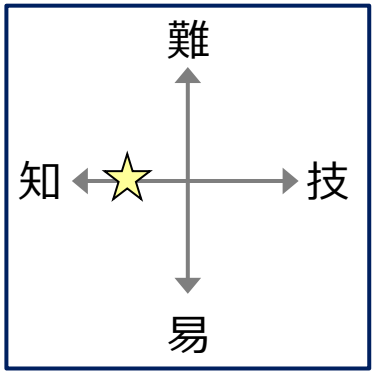
（万円）クラウドサービスの利用と労働生産性（平成28年）



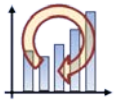
【出所】平成28年 通信利用動向調査［総務省］に基づき作成 http://www.soumu.go.jp/johotsusintokei/statistics/statistics05.html

□ クラウドサービスは小規模企業にも普及しつつあり、その利用によって労働生産性を高められる可能性があります。

総務省　ICTスキル総合習得教材



［コース3］データ分析



3-1：ビッグデータの活用と分析に至るプロセス

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_3_1.pdf

	1	2	3	4	5
［コース1］データ収集					
［コース2］データ蓄積					
［コース3］データ分析	●				
［コース4］データ利活用					

本講座の学習内容［3-1：ビッグデータの活用と分析に至るプロセス］

【講座概要】

- ビッグデータの特性の「3つのV」を説明し、それぞれの特性によって可能になる分析を示します。
- データの品質のいくつかの観点から紹介し、品質の悪いデータがもたらす社会的費用を紹介します。
- データ形式の標準化およびデータクレンジングの重要性を示します。
- 国内企業におけるデータ分析の実態、効率的なデータ分析の設計、本格的なデータ分析に至るプロセス（工程）を紹介します。

【講座構成】

座学

- [1] ビッグデータの特性と分析
- [2] データの品質と標準化・クレンジング
- [3] データ分析の設計と分析に至るプロセス

【学習のゴール】

- ✓ ビッグデータの「3つのＶ」と、各特性によって可能になる分析事例を紹介できる。
- ✓ 品質の悪いデータの社会的費用とその軽減策としてのデータ形式の標準化、データクレンジングの重要性を理解する。
- ✓ 効率的なデータ分析の設計と本格的なデータ分析に至るプロセスを説明できる。

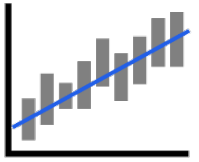
データ利用方法としての分析

◆この講座では、データの利用方法の一つとしての「データ分析」と関連事項を概説します。

- 蓄積されたデータの利用方法には大別して2種類あり、一つ**はデータベースとしての利用**、もう一つは**分析用データとしての利用**です。
  - 講座2-1で示したように「データベース」の要件として、個々のデータレコードを「検索ができること」が挙げられます。

➢ **「データベースとしての利用」**では、検索によって抽出された「個々のデータレコード」に注目します。
  - 「データベースとしての利用」は、例えば、個々のデータレコードを抽出して「カタログ、データレコード別の情報サービス」として利用できます。

➢ **「分析用データとしての利用」**では、「データ全体または一部の傾向や特徴」に注目します。
  - データの特徴や傾向を発見、把握することで、未知の情報を予測できるケースもあります。



2種類のデータの利用方法の要点・天気データでの利用例

利用方法	注目対象	利用事例	天気データでの利用例
データベースとしての利用（検索による抽出）	個々のデータレコード	カタログ、データレコードの情報利用	特定の場所、時点に関する天気情報の検索と抽出
分析用データとしての利用	データ全体または一部の傾向・特徴	傾向・特徴の発見、未知の情報の予測	天気の地域性・季節性の発見、天気予報

□ この講座では、データ分析の序論として、データの種類、品質、望ましい分析の設計を紹介します。

ビッグデータ

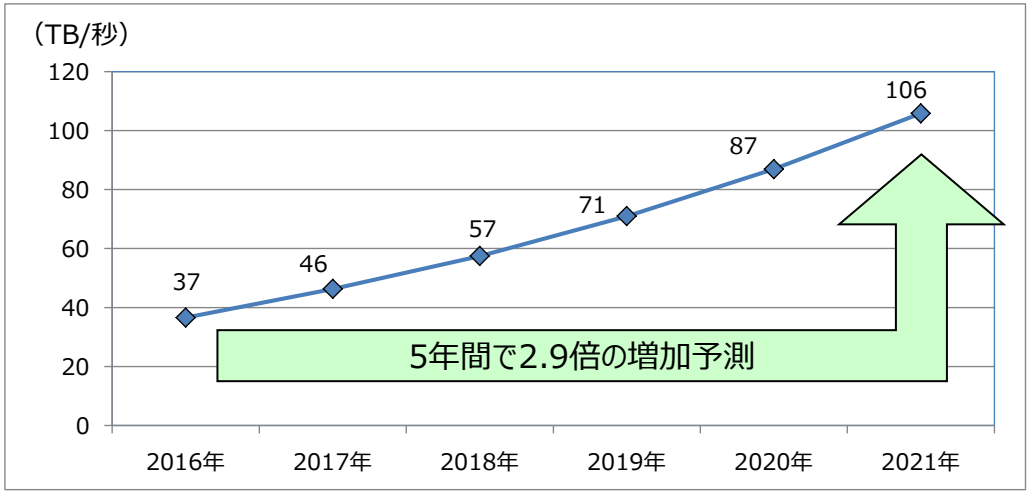
◆インターネット関連のデータは、その特性によって「ビッグデータ」と呼ばれることもあります。

- PC・スマートフォンをはじめとして、インターネットにつながる機器であるIoT機器が増加しています。
- SNS等の普及によって、一般利用者がプログラム不要で様々なデータをインターネット上に保存できるようになりました。
  - 数値データ・テキストデータのみならず、画像や動画といった容量の大きいマルチメディアデータの送受信が拡大してきています。
- 様々な機能・活動によって蓄積された膨大なデータは、その特性に応じて**ビッグデータ**と呼ばれることがあります。
- 2017年6月にシスコ社から公表された資料によれば、全世界のインターネットにおいて送受信されたデータ量は、2016年において年間1.2ZB（ゼタバイト）でしたが、2021年には3.3ZBに達すると予測されています。
  - データのサイズは、byte（バイト）から1000倍ごとにkB、MB、GB、TB、PB、EB、ZBへ単位が変わります。
  - 1年間を365日（31,536,000秒）と見なせば、年間1.2ZBのデータ通信量は、1秒間に約37TB（36,550GB）に相当します。

データサイズの単位

単位	読み方	数値表記（バイト）	
B	バイト	1	バイト
kB	キロバイト	1,000	千バイト
MB	メガバイト	1,000,000	百万バイト
GB	ギガバイト	1,000,000,000	十億バイト
TB	テラバイト	1,000,000,000,000	兆バイト
PB	ペタバイト	1,000,000,000,000,000	千兆バイト
EB	エクサバイト	1,000,000,000,000,000,000	百京バイト
ZB	ゼタバイト	1,000,000,000,000,000,000,000	十垓バイト

インターネット上で1秒間に送受信されるデータ量（予測値）



【出所】Cisco Visual Networking Index：予測と方法論［Cisco］に基づき作成
https://www.cisco.com/c/ja_jp/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf

ビッグデータの特性

◆ビッグデータが持ち得る特性として「3つのV」が挙げられます。

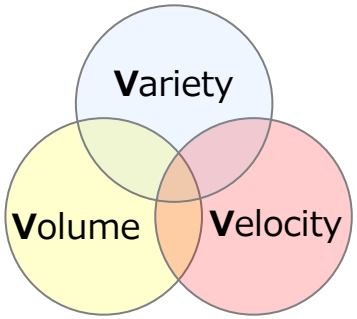
- ビッグデータの持ち得る標準的な特性としての**「3つのV」**は、2001年にアメリカのデータ分析者によって提示され、現在でもビッグデータに関する標準的な考え方となっています。

【出所】Deja VVVu: Others Claiming Gartner's Construct for Big Data［Gartner | Doug Laney］
https://blogs.gartner.com/doug-laney/deja-vvvue-others-claiming-gartners-volume-velocity-variety-construct-for-big-data/

- ビッグデータには、その特性とされる「3つのV」の**Variety（バラエティ）**、**Volume（ボリューム）**、**Velocity（ベロシティ）**のいずれかを持っていることが挙げられます。
  - 「Variety」「Volume」「Velocity」のそれぞれの頭文字の「V」から「3つのV」と呼ばれます。

ビッグデータの「3つのV」の意味

V	日本語訳	意味
Variety	データの多様性	テキスト、画像、音声といった多様な情報とファイル形式
Volume	データ量	膨大なデータ量
Velocity	データ生成速度・頻度	リアルタイムで収集できるデータ・秒単位など高頻度のデータ



- IBM社の資料では、「3つのV」に加えて、4つめのVとして「データの正確さ」の（Veracity）を挙げています。
  - 様々な組織が公表する資料によっては、「価値あるデータ」（Value）を加えて「5つのV」としているケース、10以上のVではじまる英単語を列挙しているケースもありますが、4つ以上のVの中には「3つのV」の「Variety」「Volume」「Velocity」が含まれることが標準的です。

【出所】IBM Data Engine for Hadoop and Spark（P4）［IBM］ http://www.redbooks.ibm.com/abstracts/sg248359.html

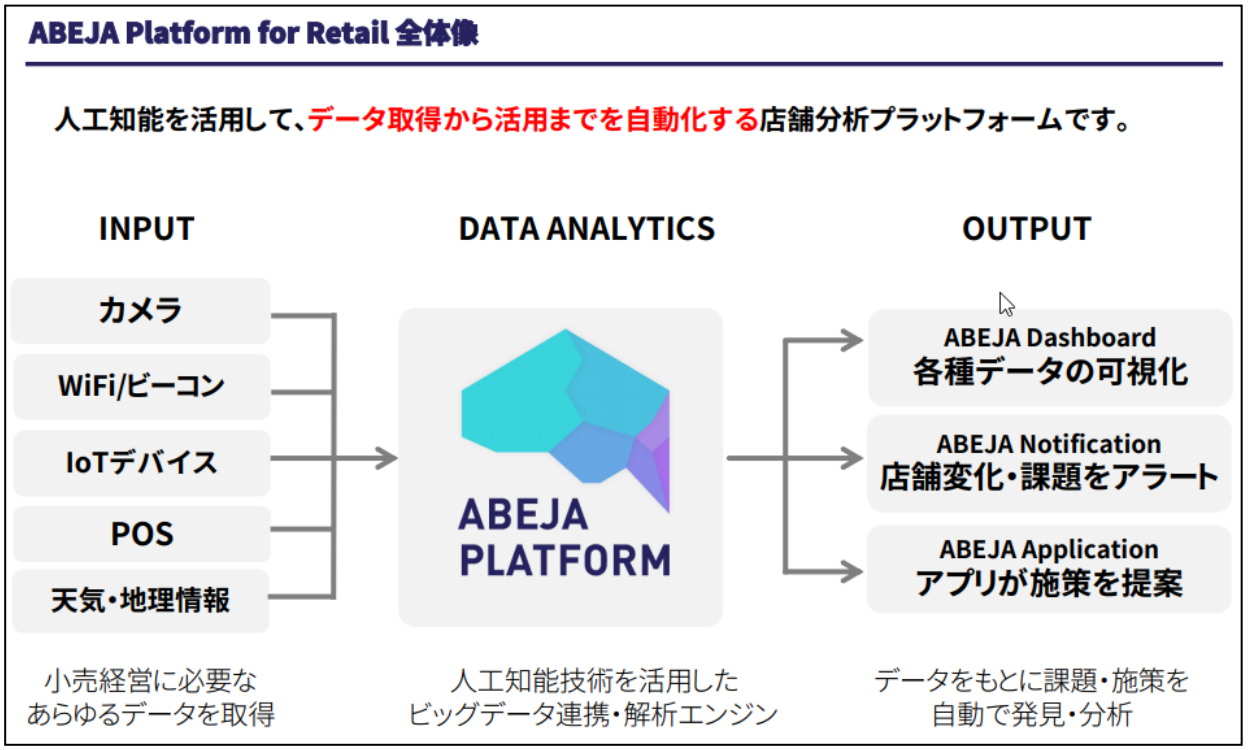
□ 3つのVはビッグデータが持ち得る特性であるため、ビッグデータであっても「対象情報やファイル形式が固定しているケース」「データ量が小さいケース」「データの生成速度が遅い、低頻度のケース」があり得ます。

Variety（多様性）により可能となる分析

◆ビッグデータのVariety（多様性）から様々なデータを統合した分析が可能となります。

- 株式会社ABEJA（アベジャ）が提供しているABEJA Platformでは小売店の店舗にカメラを設置して、来客人数をカウントし、来客者の年齢層・性別を人工知能で判定します。
- 「カメラによる画像情報」「Wi-Fi/ビーコンによる顧客の移動」「IoTデバイスによる扉の開閉状況」「POSによる売上データ」「インターネットから得られた天候情報」を組み合わせて、販売状況の管理・分析が可能となります。
  - POS（ポス）は［point of sales （system）：販売時点情報管理］の頭文字に由来するレジの販売情報管理です。

ABEJA platform for Retailにおけるカメラによる情報収集と分析概要



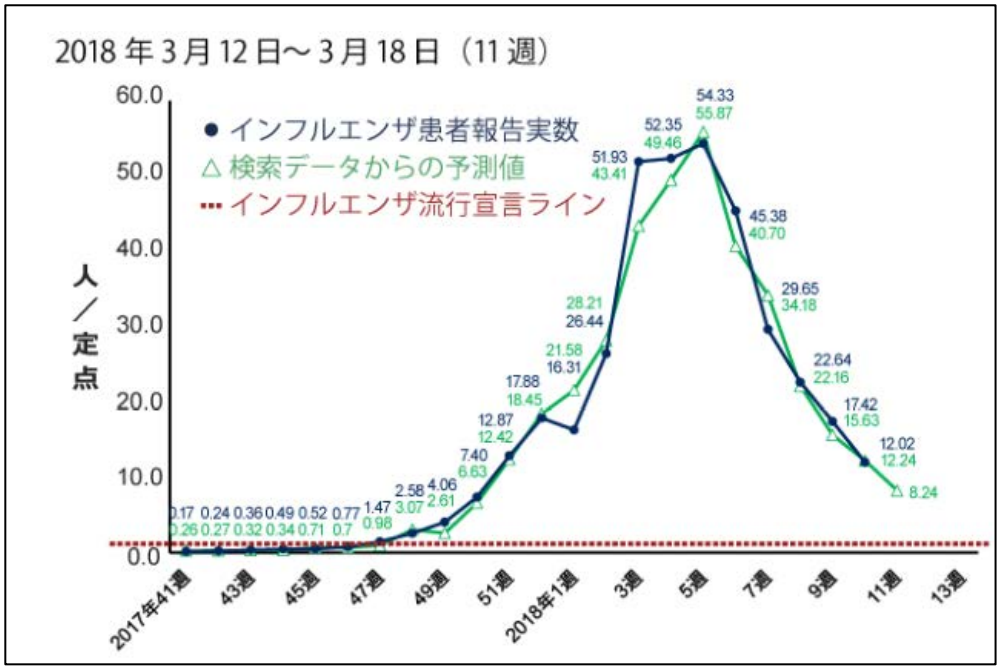
【出所】株式会社ABEJA https://abejainc.com/ja/

Volume（データ量）により可能となる分析

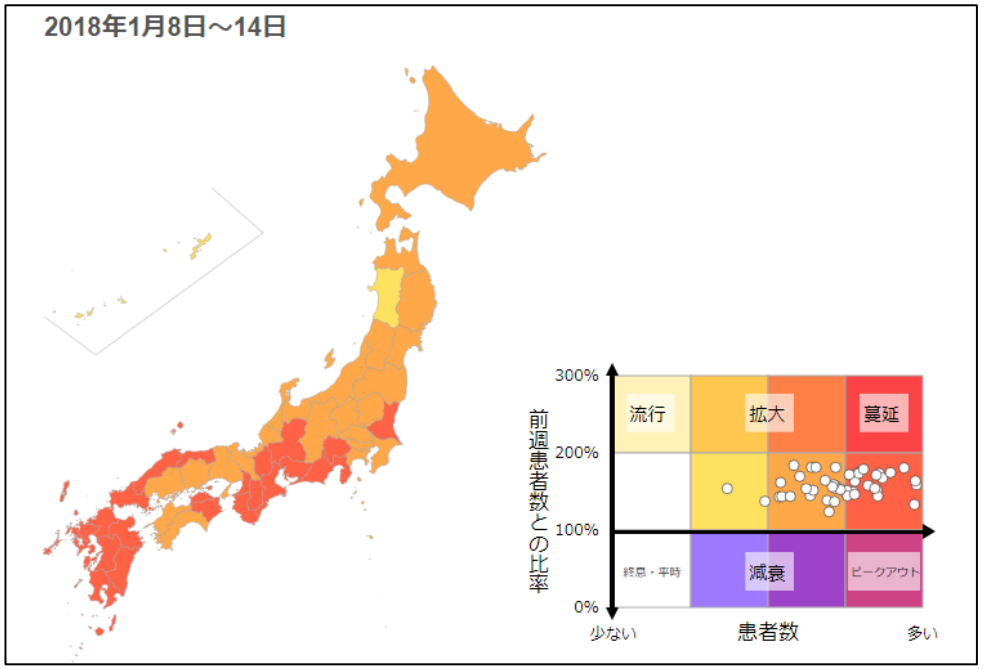
◆ビッグデータのVolume（データ量）から、膨大なデータに基づく分析が可能となります。

- Yahoo! Japanでは、時期別・都道府県別のインフルエンザの患者数と相関の高いキーワード検索数から、インフルエンザの感染数の予測値を示しています。
  - 「インフルエンザ」「発熱」「寒気」等のキーワードでの検索数を時期別・都道府県別に集計して分析用データとして活用しています。
  - 検索のキーワード、検索数と実際の患者数の対応関係を分析することで、予測の精度を一層高めることができます。
- 週単位、都道府県別にインフルエンザの「流行期」「拡大期」「蔓延期」「減衰期」の推移を確認できます。
  - ウェブサイトにおいて、データの対象期間を動かす矢印のボタンをクリックすることで、インフルエンザの流行と減衰が動的に把握できます。

検索データに基づくインフルエンザの感染予測値と患者数



インフルエンザ感染状況マップ



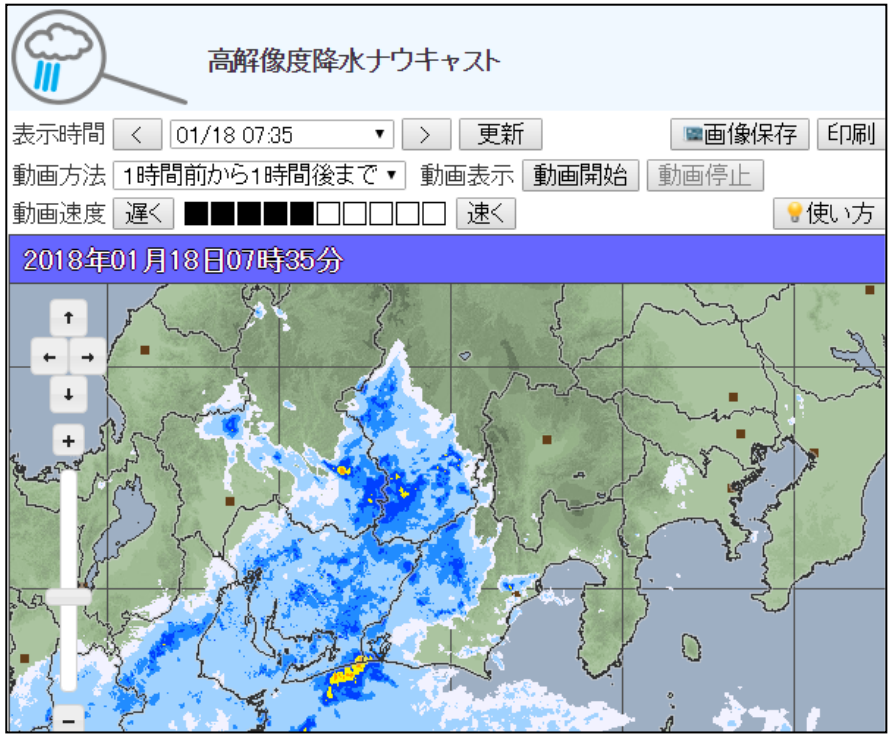
【出所】ビッグデータ分析でみるインフルエンザ感染状況：2017－2018［ヤフー株式会社］
https://about.yahoo.co.jp/info/bigdata/influenza/2017/01/

Velocity（データ生成速度・頻度）により可能となる分析

◆ビッグデータのVelocity（データ生成速度・頻度）からリアルタイムでの活用ができます。

- 気象庁が提供する「高解像度降水ナウキャスト」では、5分単位での降水状況および、1時間後までの降水予想を地図上に示します。
  - 詳細な地図で表示できるため、ゲリラ豪雨の予測にも利用できます。

高解像度降水ナウキャスト（気象庁）



【出所】高解像度降水ナウキャスト［気象庁］
https://www.jma.go.jp/jp/highresorad/

- ネットショッピングサイトのAmazon.co.jpでは、各ユーザーの購入予定の商品に合わせて、即座にお勧め商品を表示します。
  - 講座3-5に示す「アソシエーション分析」に基づいて表示します。

購入予定商品に合わせた商品推薦（Amazon.co.jp）

【出所】Amazon.co.jp　https://www.amazon.co.jp/

構造化データ、半構造化データ、非構造化データ

◆ビッグデータは、人間にとって読みやすく、分析しやすい構造化データだけではありません。

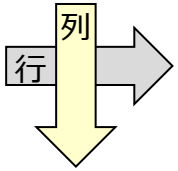
- ビッグデータはその特性である**多様性（Variety）**から**構造化データ**、**非構造化データ**のケースもあります。
  - 講座2-1のデータベースの説明においても、「構造化データ」「半構造化データ」「非構造化データ」を紹介しました。
  - 半構造化データの「XML」「JSON」に関しては、講座1-4のAPIで利用されるファイル形式として説明しました。

「構造化データ」「半構造化データ」「非構造化データ」に関する説明表

データ種別	説明	データ形式の例
構造化データ	二次元の表形式になっているか、データの一部を見ただけで二次元の表形式への変換可能性、変換方法が分かるデータ	CSV、固定長、Excel（リレーショナルデータベース型）
半構造化データ	データ内に規則性に関する区切りはあるものの、データの一部を見ただけでは、二次元の表形式への変換可能性・変換方法が分からないデータ	XML、JSON
非構造化データ	データ内に規則性に関する区切りがなく、データ（の一部）を見ただけで、二次元の表形式に変換できないことが分かるデータ	規則性に関する区切りのないテキスト、PDF、音声、画像、動画

二次元の表形式の構造化データ

世帯名	大人1	大人2	子供1
山田家	世帯主	妻	長女



XML形式の半構造化データ

```
<世帯>
  <世帯名>山田家</世帯名>
  <大人>世帯主</大人>
  <大人>妻</大人>
  <子供>長女</子供>
</世帯>
```

画像形式の非構造化データ

- 一般に半構造化データ、非構造化データは、分析を行う前にデータ整理や変換が必要です。

日本政府の構造化・非構造化データの提供サイト

◆日本政府のウェブサイトには、公的統計の構造化データを提供するe-Stat、非構造化データを含めて幅広く提供するDATA.GO.JPがあります。

- 公的統計の調査結果データを提供しているe-StatではExcel形式、CSV形式のデータをダウンロードでき、構造化データを提供しているウェブサイトと言えます。
  - 講座4-3の参考2にて、Rによる利用方法を紹介するe-Stat APIにおいては、e-Statが蓄積するデータを半構造化データ（XML、JSON）の形式でも提供しています。
- DATA.GO.JP(データカタログサイト)は、「政府の報告書などのPDF」「政府ウェブサイトのHTML」「報告書内の画像JPEG」といった非構造化データを含めて幅広く提供しています。
  - DATA.GO.JP(データカタログサイト)は、講座4-1でも紹介する日本政府のオープンデータの提供サイトでもあります。

公的統計の構造化データを提供するe-Stat

【出所】e-Stat［総務省］https://www.e-stat.go.jp/

非構造化データを含めて提供するDATA.GO.JP

【出所】DATA.GO.JP［総務省］http://www.data.go.jp/

データの品質

◆データには品質があり、データの品質が悪ければ、利用や分析における障害となります。

- 構造化データに限っても、重複するデータ、表記揺れ等があり、**データの品質が悪いケース**があります。
- 国際データマネジメント協会の英国支部の資料では、**データの品質**には6つの主要基準があると示しています。
  - このデータの品質基準には、客観的でデータ固有の基準のみではなく、利用者の主観的な有用度合いに依存する「Timeliness（適時性）」、他のデータとの照合しやすさとして「Consistency（一貫性）」が含まれていることが特徴的です。

DAMA UKのレポートによるデータの品質に関する6つの主要基準

基準	説明	品質が損なわれている例
Completeness（網羅性）	保存されているデータの割合は、潜在的な全データに対して「100%網羅」していること	部分的なデータ
Uniqueness（唯一性）	特定された対象が、2行以上にわたって記録されていないこと	重複するデータレコード
Timeliness（適時性）	要求する時点の現実を表している程度	速報性がない調査データ、低頻度の調査データ【利用者のニーズに依存】
Validity（正当性）	定義されている構文規則（フォーマット、型、範囲）に正しく準拠していること	表記揺れ、誤記入、数値が入るべきデータ項目へのテキストの記入
Accuracy（正確性）	記述している現実世界の対象やイベントを正確に表している程度	測定誤差の大きいレコード
Consistency（一貫性）	データセット内、データセット間で一つの定義に対して、複数の表現等の相異がないこと	データセット間の「西暦と和暦」の混在【他のデータセットとの関係に依存】

【出所】THE SIX PRIMARY DIMENSIONS FOR DATA QUALITY ASSESSMENT［DAMA UK］
http://www.damauk.org/RWFilePub.php?&cat=403&dx=1&ob=3&rpn=catviewleafpublic403&id=106193

- データの品質が悪ければ、データ利用・データ分析ができなかったり、誤った出力が得られたりします。
  - ある側面でデータの品質が悪かったとしても、利用目的によっては不都合がないケース、データクレンジングによって修正ができるケースもあります。

品質の悪いデータによる社会的費用

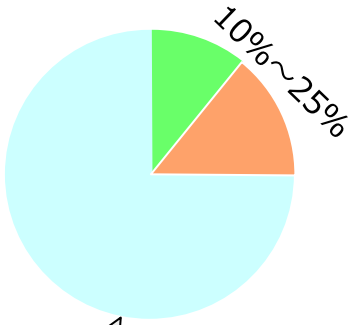
◆品質の悪いデータは、大きな社会的費用を生んでいます。

- 2016年にIBM社より公刊された書籍では、「品質の悪いデータがアメリカ経済に与えているコスト推定値は年間3.1兆ドル」と紹介しています。

【出所】Data Engine for Hadoop and Spark（P4）［IBM］
Http://www.redbooks.ibm.com/abstracts/sg248359.html

- 品質の悪いデータが生み出す社会的費用は、「正しいデータが確認できないことによる機会損失」「データの廃棄や追加的な作業によるコスト」「不正確なデータ利用に基づくコスト」が挙げられます。
  - 社会的費用には実際に支出した費用のみならず、データの品質が悪いことによって得られなかった潜在的な利益も含まれます。

アメリカにおける「品質が悪いデータが生み出すコスト」に関するインフォグラフィック（翻訳）



品質の悪いデータがもたらす経営上のコストは、組織の収益の10～25%にのぼる。

ヘルスケア部門における質の悪いデータによる費用は3140億ドル（→31兆4000億円）となっている。

質の悪いデータによるアメリカ経済へのコストは3兆ドル（→300兆円）であり、2011年の財政赤字の2倍以上となっている。

正しいデータが分からないことで、ある主要な小売業者が被った被害は年間300万ドル（→3億円）となっている。

典型的なIT予算の50%は、使えない情報とその補修に費やされる。

平均的な会社は、不正確なデータのために想定した受取人に届かない郵送料に毎年18万ドル（1800万円）を費やしている。

＊ オリジナルのインフォグラフィックの値から「1ドル＝100円」にて換算

【出所】SOFTWARE AGのインフォグラフィックに基づき作成　https://lemonly.com/work/the-cost-of-bad-data

データ形式の標準化とデータクレンジングの重要性

◆「データ形式の標準化」や「データクレンジング」によってデータの品質を高めることができます。

- 2015年に総務省 統計委員会から公表された報告書では、**ビッグデータ活用における課題**として、「**データクレンジング技術の高度化**、企業・業界横断的に**データ形式の標準化**」を挙げています。
  - 「データ形式の標準化」や「データのクレンジング」によって、品質の悪いデータによる社会的費用を軽減することができます。

【出所】公的統計におけるビッグ・データの活用に関する調査研究［［総務省（調査委託先：株式会社 NTTデータ経営研究所）］
http://www.soumu.go.jp/main_content/000422923.pdf

- **データ形式の標準化**は、公的機関や業界等のコンソーシアムが形式を定め、データの提供者が実施する根本療法に相当し、**データクレンジング**は、一般に分析者・利用者自身が行う対処療法に相当します。
  - 根本療法としての「データの標準化」の推進が重要である一方で、対処療法としての「データクレンジング」の技術が必要なケースもあります。

ビッグデータ活用における課題（品質の悪いデータに対する対応策）

対応策	主な実施主体	意味
データ形式の標準化	データ形式の決定：公的機関・業界等のコンソーシアム 標準化の実施：データ提供者	定められた基準によって、データのファイル形式や変数名を統一し、利用可能なデータレコードを抽出することによって、データの利用やデータセット同士の連結を容易にすること
データクレンジング	分析者・利用者	データレコードの重複、データ内の誤記、表記の揺れなどを修正・統一することでデータの品質を高めること

- 品質の良いデータであっても、利活用に適する形への「データ整理・抽出」や「データ加工・結合」は必要であり、「データクレンジング」「データ抽出・加工・結合」との技術は、データ分析者・利用者にとって重要です。
  - 「データクレンジング」「データ抽出・加工・結合」といった分析前の一連の作業は、「データの前処理（まえしょり）」とも言われます。

データ形式の標準化政策

◆日本政府では「データ形式の標準化」に関する政策を推進しています。

- 総務省の自治体クラウドポータルサイトでは、地方自治体が保有するデータの標準化を推進するべく中間標準化レイアウト仕様を公開してます。
  - 中間標準レイアウト仕様では「住民基本台帳」「印鑑登録」「戸籍」といった行政書類の標準フォーマットを公開しています。
  - 地方自治体のデータ形式を標準化することで、広域でのデータ連携、住民の転居に伴うデータの移行をスムーズに行うことができます。

- 経済産業省では、消費・購買データの標準的なフォーマットを設定し、電子化された買物レシート（電子レシート）の標準仕様を検証する実証実験を2018年2月に実施しました。
  - レシートのデータを電子化・標準化することで、様々な商店・ネットショッピングサイトの消費・購買データを一括して取り扱うことができるようになります。

総務省（自治体クラウドポータルサイト）の中間標準レイアウト

○中間標準レイアウト仕様V24

・(表形式)一括ダウンロード(zip)(12.9MB)

・(XML形式)一括ダウンロード(zip)(8.4MB)

	一括ダウンロード(zip)	表形式					XML形式		改訂履歴
		移行ファイル構成表	移行ファイル関連図	データ項目一覧表	コード構成表	コード一覧	レイアウト仕様	サンプル	
1. 住民基本台帳									
2. 印鑑登録									
3. 住登外管理									
4. 戸籍									
5. 就学									
6. 選挙人名簿管理									

【出所】中間標準レイアウト仕様［総務省］
http://www.soumu.go.jp/main_sosiki/jichi_gyousei/c-gyousei/lg-cloud/02kiban07_03000024.html

経済産業省の電子レシート実証実験用アプリ

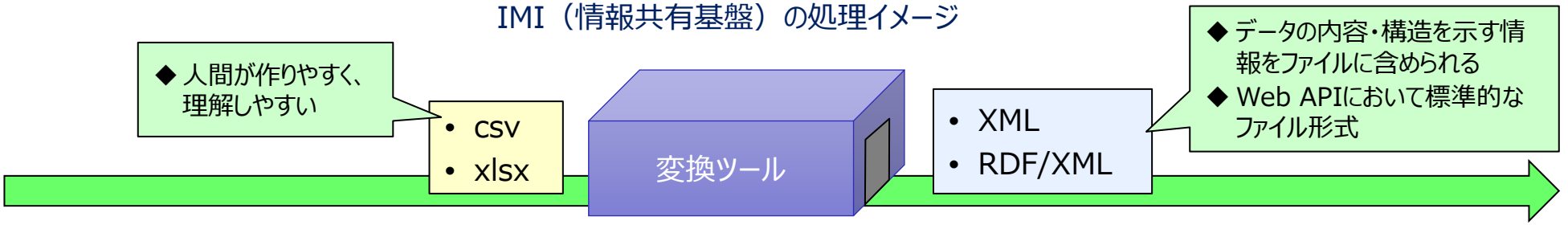
【出所】電子レシートの標準仕様を検証する実験を行います［経済産業省］
http://www.meti.go.jp/press/2017/01/20180131004/20180131004.html

データ形式の標準化ツール

◆日本政府ではデータ形式を標準化するツールの開発・公開を行っています。

- 経済産業省が設置し、情報処理推進機構（IPA）が事務局を担当する**IMI（情報共有基盤）**では**「DMD Editor」というデータ形式の標準化・変換を行うウェブツール**を提供しています。
  - IMIは［Infrastructure for Multilayer Interoperability（情報共有基盤）］の略であり、DMDは［Data Model Description（データモデル記述）］の略となっています。
  - DMD Editorはウェブサイトに「csv」や「xlsx」をアップロードすることで、自動で「RDF/XML」「JSON」といったファイルに変換できます。
  - 「RDF/XML」は講座1-5にて示したXMLに外部からの特定・リンクを可能とした規格であり、講座4-1の「機械判読への適性の5段階」でも紹介します。

IMI（情報共有基盤）の処理イメージ



【出所】IMI（情報共有基盤）https://imi.go.jp/

災害被災者支援 支援物資提供データ（二次元の表形式）

場所	提供者	支援物資	提供状況
○×小学校	NPO◆◆	飲料水	提供中
△□公民館	株式会社■■	米	提供準備中

RDF/XML

```
…<ic:場所 rdf:resource="○×小学校"/>
   <ic:状況型>
     <rdf:Description>
       <ic:説明-単純型>飲料水</ic:説明-単純型>…
```

JSON

```
…{"@id":"_:b2",
"http://imi.go.jp/ns/core/rdf#説明-単純型":
[{"@value":"飲料水"}],
"http://imi.go.jp/ns/core/rdf#名称型":
[{"@id":"_:b4"}]},...
```

- 2018年1月決定の「デジタル・ガバメント実行計画」において、**日本政府はIMIを整備・活用**する旨が示されました。

【出所】デジタル・ガバメント実行計画［eガバメント閣僚会議］https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/egov_actionplan.pdf

データのクレンジングツール

◆無償利用可能なデータクレンジングツールもありますが、日本語への対応は不十分です。

- データクレンジングを行うための無償利用が可能な英語版ソフトウェアとしてOpenRefineが挙げられます。

【出所】OpenRefine http://openrefine.org/

- 日本語は英語に比べても、漢字表記や送り仮名の違い等の表記揺れが多く、標準化（名寄せ）は、より重要です。

住所表記・会社表記のデータ形式の標準化（名寄せ）例

住所の表記揺れ	
霞ヶ関1丁目1番地	霞が関1丁目1番地
霞ヶ関1丁目1	霞が関1丁目1
霞ヶ関1-1	霞が関1-1

住所表記の標準化の取り組み例

- 標準記載法の策定と公表
- 標準記載名データベースの公表
- 表記揺れの統一エンジンの公開

住所表記の標準化例

霞が関１丁目１
1-1 Kasumigaseki
〒100-0013
緯度: 35.675836 経度: 139.754734

- 住所の表記においては、「ヶ」と「が」の混在、丁番地の表記が不統一となっている事だけでも、一貫性が損なわれてしまいます。

ソニー株式会社の表記揺れ				
ソニー株式会社	Sony株式会社	SONY株式会社	Ｓｏｎｙ株式会社	ＳＯＮＹ株式会社
ソニー（株）	Sony（株）	SONY（株）	Ｓｏｎｙ（株）	ＳＯＮＹ（株）
ソニー(株)	Sony(株)	SONY(株)	Ｓｏｎｙ(株)	ＳＯＮＹ(株)
ソニー㈱	Sony㈱	SONY㈱	Ｓｏｎｙ㈱	ＳＯＮＹ㈱

会社表記の標準化例

ソニー株式会社
Sony Corporation
東証一部 6758（電気機器）
設立年月日　1946年5月7日

- 日本人が見れば、上記16種の企業表記は同一の企業だと分かりますが、文字列が異なるためデータ集計時には異なる企業として扱われてしまいます。
- 法人マイナンバー（法人番号）を利用すれば、正式な企業名を確認することができ、同じ企業名が複数ある場合でも企業を特定することができます。

- 日本語のデータクレンジングは、個々のケースに合わせてExcelやプログラミングで行っているケースが多くなっています。

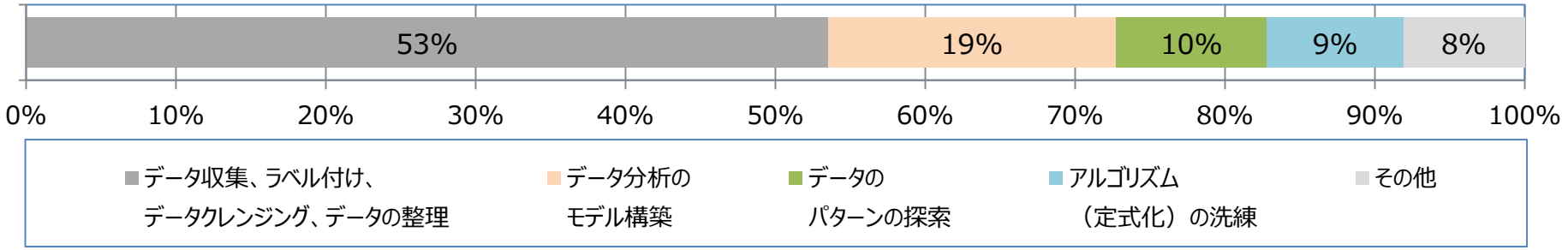
□ 講座3-2では、Excelを用いて日本語の表記揺れの統一を含むデータクレンジングの実習を行います。

データクレンジングの負担

◆データ分析において、データクレンジング・データ整理は時間がかかり、好まれない作業です。

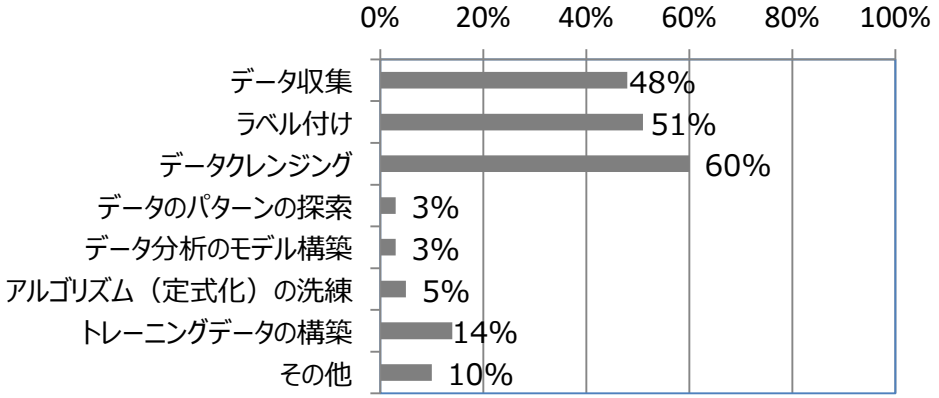
- 2017年2月における世界のデータサイエンティスト（データ分析者）197名に対する調査では、データクレンジングを含むデータの前処理が最も時間を割いている業務と回答した者が過半の53%となっています。

データサイエンティストがデータサイエンスの業務時間で最も時間を割いている業務

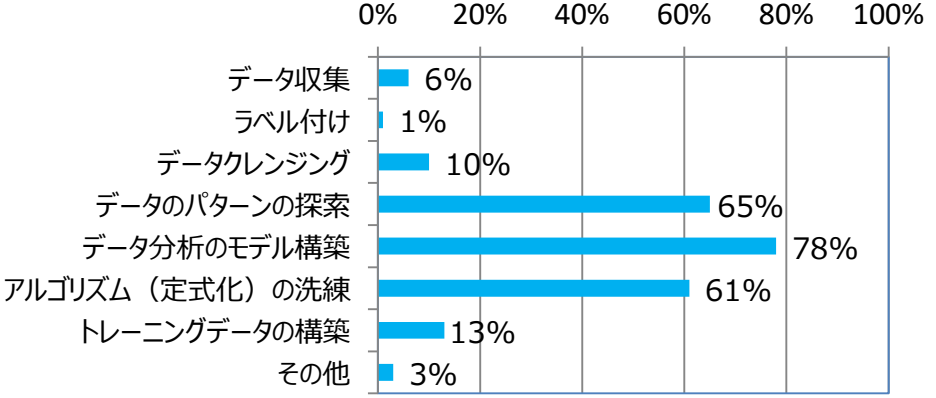


- データクレンジングは、データサイエンティストの業務の中で「最も楽しめない業務」として挙げられています。
  - データサイエンティストが最も楽しめる業務として「データ分析のモデル構築」が挙げられています。

データサイエンティストが最も楽しめない業務（最大3つ）



データサイエンティストが最も楽しめる業務（最大3つ）



【出所】2017 Data Scientist Report［CrowdFlower］に基づいて作成

□ データクレンジングは労働時間の大部分を占めるとともに心理的な負担になっており、その軽減が課題となっています。

国内企業におけるデータ分析の実態

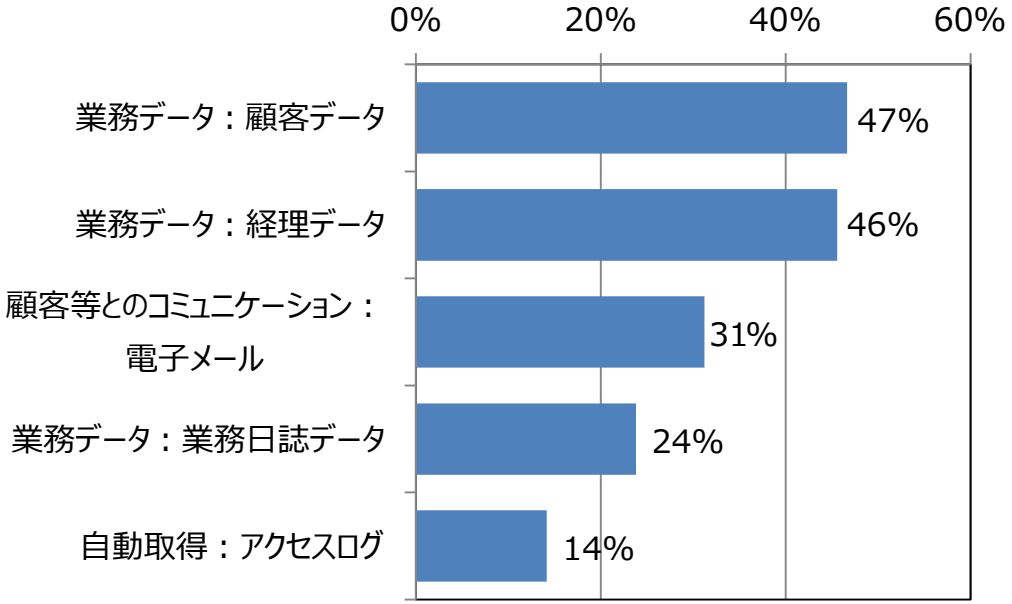
◆国内企業では「業務データ」を「見える化」するデータ分析の方法が、最も多くなっています。

- 総務省の2014年度の委託調査によれば、国内企業4,672社のうち72%の3,357社がデータ分析を行っています。
  - 本スライド下部の2種類のグラフはいずれも、データ分析を行っている3,357社が分母となっています。

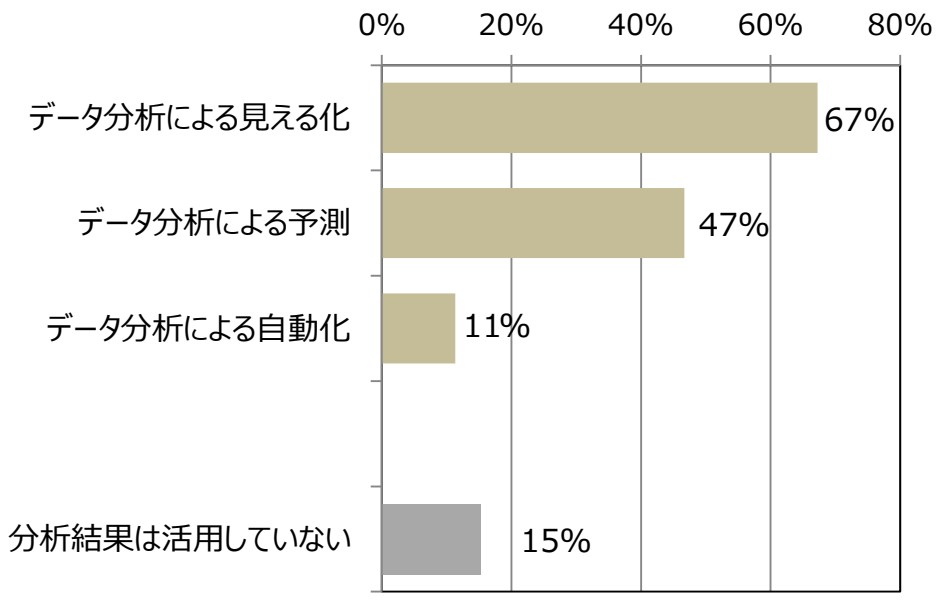
- 分析に活用しているデータとして「顧客データ」、「経理データ」の割合が高くなっています。
  - いずれも意図的に取得したデータではなく、自然に集まる業務データとなっています。

- データ分析の活用方法として、最も割合が高いのは「データ分析による見える化（可視化）」の67%です。
  - 「見える化（可視化）」とは、図表作成などを行うことでデータを分かりやすく示すことを指しています。

分析に活用しているデータの割合（複数回答：降順上位5位）



データ分析の活用方法（複数回答）



【出所】ビッグデータの流通量の推計及びビッグデータの活用実態に関する調査研究［総務省（調査委託先：株式会社 情報通信総合研究所）］に基づき作成
http://www.soumu.go.jp/johotsusintokei/linkdata/h27_03_houkoku.pdf

□ 自然に集まる業務データを活用し、見える化（可視化）して、分かりやすく表すことが分析の第一歩となっています。

より良いデータ分析の設計

◆より良いデータ分析の設計として、目的や分析課題を明確にすることが挙げられます。

- 私達はビジネスにおいても、私生活においても、様々な目的があり、それに対する意思決定（選択）をしています。
- データ分析を行うことで、目的に対して、より効果的な意思決定（選択）を行うことができます。
  - 必ずしも自分自身でデータ分析を行う必要はなく、データ分析を依頼することも、公表されている分析結果のみを確認することもあります。
  - データ分析を行わない人や場合においても、まずは定量的なデータや指標を確認する姿勢が重要です。

ビジネスの目的例：売上総額を上げたい

- ◆ 売上総額は［販売単価］×［販売個数］で構成されている。
- ◆ ［販売単価］は企業が決められるが、［販売単価］を上げれば［販売個数］は下がる関係にある。

⇨ **データ分析によって、売上総額を最大化するための［販売単価］を知りたい。**

私生活の目的例：ダイエット（減量）したい

- ◆ ダイエットには［食事制限］と［運動］の両方に効果があるとされている。
- ◆ ［食事制限］と［運動］をどのように組み合わせることが、ダイエットに効果的かが分からない。

⇨ **データ分析によって、ダイエットに効果的な［食事制限］と［運動］の組み合わせを知りたい。**

- あらかじめ**「何をしたいのか？（⇒目的）」**や**「何を知りたいのか？（⇒分析課題）」を明確にすること**で、意思決定（選択）に反映できるデータ分析の方針を定められるとともに、効率的に分析作業ができます。
  - データが手元にありつつも、データ分析の目的や分析課題を明確にしにくいケースにおいては、見える化（可視化）によってデータをく図表に表し、実態や外れ値を確認することで、高度な分析へのヒントが得られるケースもあります。

本格的なデータ分析に至るプロセス（工程）

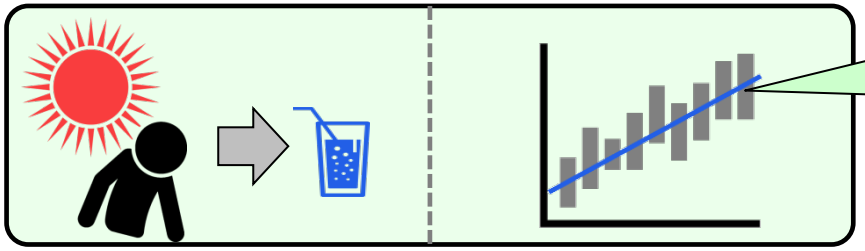
◆本格的なデータ分析に至る前には、いくつかのプロセス（工程）があります。

- データ分析を行う際の目的や分析課題には、様々なケースが考えられます。

目的・分析課題の設定例

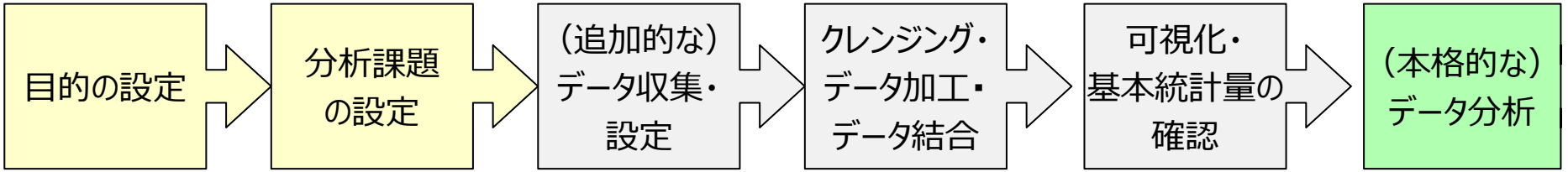
目的の設定	分析課題の設定
最適な仕入れ量の設定	環境と売上の関係を定量的に測定
購入機材の選択	各機材の費用対効果を測定
生産費用の削減	生産機械の最適なメンテナンス時期の把握
顧客満足度の向上	顧客満足度が増減する要因の特定

- 「定性的には当たり前のこと」であっても、定量的な関係な測定が分析課題となることもあります。
  - 「気温が上がれば、冷たい飲み物の販売量が増加する」ことは、感覚的・定性的に当たり前ですが、「気温1度の上昇につき平均○本の増加」「気温△度において、平均□本の販売量」という定量的な関係はデータ分析を行わないと把握できません。



【分析結果例】
◆ 気温1度の上昇につき平均12本の販売量増加
◆ 気温30度の平均販売量は123本

- 本格的なデータ分析の前には、「目的の設定」「分析課題の設定」に続く一般的なプロセス（工程）があります。



「目的・課題」に基づく「(追加的な）データ収集・設定」

◆「目的、課題ありき」の分析では、追加的なデータ収集や仮設定を検討することができます。

- 目的や分析課題が明確になっている「目的・分析課題ありきの方法」では、手元にないデータ項目があっても、追加的な収集や外部からの提供を検討することができます。
  - 追加データを入手するには、費用や労力がかかるため、データ取得自体の費用対効果も検討する必要があります。

「目的・分析課題ありきの方法」と「手元にあるデータからできることを考える方法」の比較

◆ 目的・分析課題ありきの方法 ⇒
- ○ 効率的な分析作業となりやすい
- ○ ストーリー・実益のあるレポートとなりやすい
- ○ **不足データの入手や設定を検討できる**

ゴール

◆ 手元にあるデータからできることを考える方法 ⇒
- × 効率的な分析作業となりにくい
- × ストーリー・実益のないデータ集になりやすい
- × **不足データは検討・言及できない**

スタート

- 入手できないデータ項目がある場合でも、近似値や仮定による設定を与えて分析をするケースもあります。
  - 利用可能なデータが利用したいと時点や地点と乖離しているなど、近似値のデータしか入手できないケースもあり得ます。
- 近似値のデータしか利用できないなど、データの品質が悪いケースでも、利用データの注意点を記載すれば、分析レポートとして提出・公表することができます。
  - 実際のデータ分析においては、万全の品質のデータが揃っていることは稀です。品質の悪いデータを利用しても、分析結果には大きな影響がないケース、品質の良いデータの収集のきっかけになるケースもあります。

データクレンジング・データ加工・データセットの結合

◆必要に応じて、データクレンジング、データ加工、データセットの結合を行います。

- 重複レコードや表記揺れがあれば、それらを修正・補正する**データクレンジング**を行います。
- 分析対象データの抽出や生年データから年齢データへ変換するなど、必要な**データ加工**があれば行います。
- 構造化データにおける**データセットの結合**は、同種のデータを追加し、**行（データレコード）が増加するアペンド（append）型**と外部データとの照合などによって**列（変数）を追加するマージ（merge）型**に分かれます。
  - **アペンド型のデータ結合**は、特定の県のデータに、比較対象としての他県のデータを追加するなど、**同じ変数でデータレコードを追加し、比較する範囲を広げるデータ結合**です。
  - **マージ型のデータ結合**は、特定の県の人口のデータに、可住地面積のデータを加えるなど、**新たな変数を追加し、新しい視点を与えるデータ結合**です。

データ結合の事例（アペンド型・マージ型）

都道府県	人口（万人）
神奈川県	913
静岡県	370

アペンド型のデータ結合

都道府県	人口（万人）
神奈川県	913
静岡県	370
愛知県	748

行（データレコード）の追加

マージ型のデータ結合

都道府県	人口（万人）	可住地面積（km²）
神奈川県	913	1,471
静岡県	370	2,749

列（変数）の追加

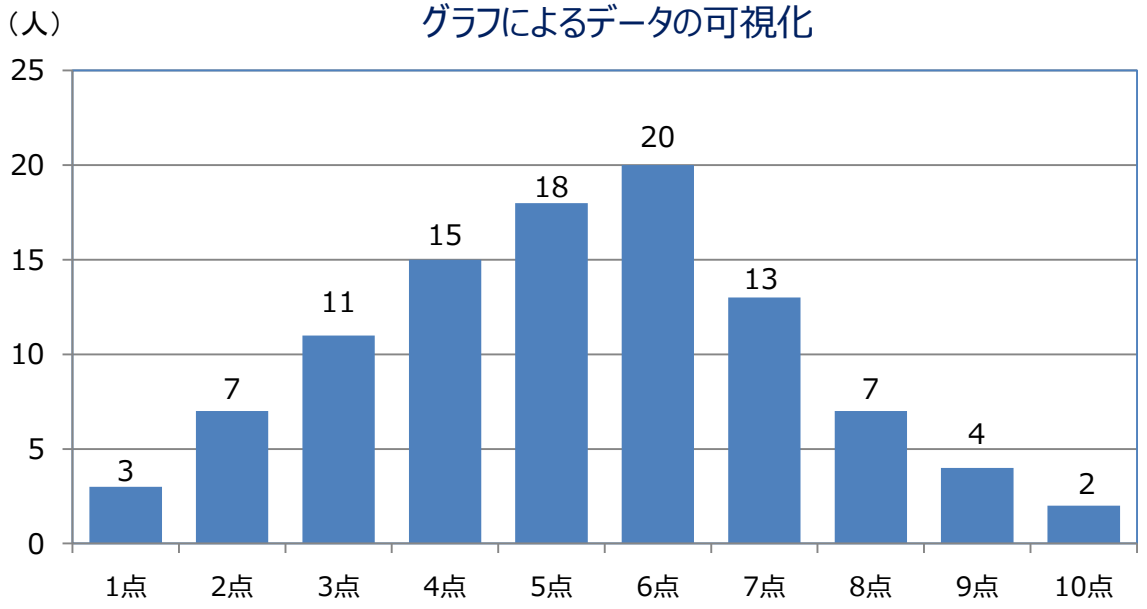
□ ExcelのVLOOKUP関数を利用したマージ型のデータセットの結合は、講座3-2にて紹介します。

データの可視化、基本統計量の導出

◆データの可視化や基本統計量を導出することで、データの全体像および外れ値を確認します。

- 本格的なデータの分析を行う前に、グラフ等で視覚的にデータの状態を確認する**データの可視化**によって、データの全体像や外れ値を把握します。
  - 突出した外れ値は、観測エラーや記入ミスとして除外すべきケースもあれば、現実の突出した値を示し、価値ある分析の糸口となるケースもあります。
- 各変数の代表的な値、バラツキに関する指標、最大値、最小値などの**基本統計量**を算出し、データの特徴を概観します。

グラフによるデータの可視化

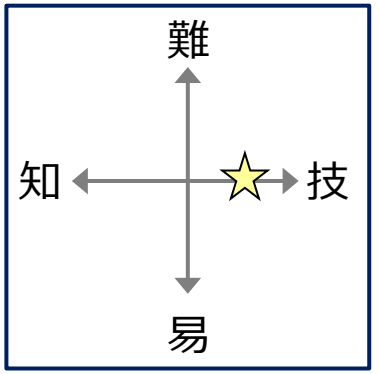


基本統計量の導出

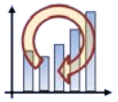
基本統計量	
平均値	5.23
最頻値	6
第1四分位	4
中央値（第2四分位）	5
第3四分位	7
最小値	1
最大値	10
分散	4.18
標準偏差	2.04

□ Excelにおけるデータの可視化については講座3-2、基本統計量の導出については講座3-3で説明します。

総務省　ICTスキル総合習得教材



［コース3］データ分析



3-2：データのクレンジングと可視化

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_3_2.pdf

	1	2	3	4	5
［コース1］データ収集					
［コース2］データ蓄積					
［コース3］データ分析		●			
［コース4］データ利活用					

本講座の学習内容［3-2：データのクレンジングと可視化］

【講座概要】

- Excelのフィルター、ステータスバーを利用したデータチェックの方法を紹介します。
- Excel関数を利用した基本的なデータクレンジングの方法を説明します。
- Excel関数を利用したデータの整理、データセットの結合、データ集計の方法を示します。
- Excelのグラフ作成による可視化とグラフの使い分けを紹介します。

【講座構成】

実習

- [1] Excelにおけるデータチェック
- [2] Excel関数によるデータクレンジング
- [3] Excelにおける分析用データ確認と抽出
- [4] Excelにおけるデータセットの結合と集計
- [5] Excelにおけるデータの可視化

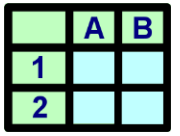
【学習のゴール】

- ✓ Excelのフィルター、ステータスバーを利用して、表記揺れ、異常値のチェックができる。
- ✓ Excel関数を利用して基本的なデータクレンジングができる。
- ✓ Excel関数を利用してデータ整理、集計ができる。
- ✓ Excelのグラフ作成で基本的な可視化ができる。

Excelによるデータクレンジング

◆Microsoft Excelを利用したデータクレンジングの基本操作を説明します。

- **データクレンジング**とは、**分析の障害となる異常値、重複データ等を取り除き、分析しやすい状態にすること**です。
  - 講座3-1でも紹介したように、データクレンジングにかかる時間が本格的な分析作業以上の時間となることがたびたびあります。データクレンジングを効率的に行うことはデータ分析、活用において重要です。
  - データクレンジングの必要はないデータであっても、本格的な分析前のデータ整理をはじめとして、より広義の「データの前処理」が必要となります。
- 本講座では、構造化（表形式）データのクレンジングの方法を紹介します。
  - 半構造化データ・非構造化データを構造化データへ変換するデータクレンジング・整理もあります。



- 一般に普及しているMicrosoft Excelを用いて、プログラミング不要で行えるデータクレンジングを紹介します。
  - Excelによるプログラミング不要のデータクレンジングは、技術面とコンピュータ環境面の制約が少ないため、組織・チーム内で依頼することが容易です。
  - 操作を例示したスクリーンキャプチャはMicrosoft Excel 2010で示していますが、Excel 2010以降であれば概ね同様の操作が実行できます。

【Excelにおけるデータクレンジングについて】

◆ Excelによるデータクレンジングには、ソフトウェアが広く普及しており、視覚に基づく直感的な操作がしやすいという長所がある一方で、作業プロセスとなるプログラムコードや作業記録となるログが自動で残らない短所があります。

この講座では、作業記録が残りやすいExcel関数を使ったデータクレンジングを説明します。

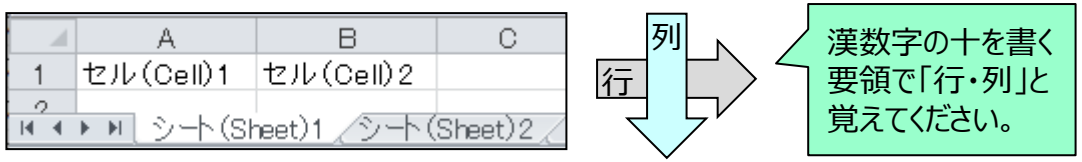
- データクレンジングの記録を残しておくことで、再度同じクレンジングを行う場合、途中からクレンジングの方法を変える場合に便利です。
- データ分析においては、他の人が行っても同じ分析結果を導出できる「客観性」や「再現可能性」が重要です。自分自身が理解するのみならず、他の人に説明できるように、他の人でも同じデータクレンジングが行えるように記録することが必要となります。
- 講座4-3で紹介するようにプログラミング言語を使ったデータ分析は意識せずとも、プログラムコードやログを残すことができます。

■ この講座では「【実習用データ】ICT3-2_データクレンジングと可視化.xlsx」を用いて実習を行います。
利用するExcelのシート番号は、各スライド右上の〔〕内に示します。 http://www.soumu.go.jp/ict_skill/dc/ict_3_2data.zip

通し番号の挿入

◆Excelにおけるデータクレンジングおよびデータ整理には通し番号を作っておくと便利です。

- Excelでは電子ファイル全体を**ブック**、ブック内の区切られたページを**シート**、各シート内の入力欄の枠を**セル**といいます。
  - Excelでは「A1」や「C3」と「列のアルファベット・行の番号」の組み合わせで表現されるセルの位置を「セルの番地（cell address）」や「セル番号」といいます。
- Excelでは水平線側を**行**、垂直線側を**列**と呼びます。
  - 講座2-1でも紹介したリレーショナルデータベースと呼ばれる構造化データの格納に適したデータベースにおいても、水平線を「行」、垂直線を「列」といいます。



- この講座では事例として、シート〔1〕にある文房具店の売上を示す構造化データをExcelでクレンジングするケースを考えます。
  - 構造化データ全体を「データセット」、行毎の個別の売上情報を「データレコード」と呼びます。
- 実習用データのシート〔1〕のように、元のデータセットに通し番号がない場合は、一番左に**［通し番号］**の列を作っておきます。
  - ［通し番号］は、行番号のIDとしても利用でき、データセットの全レコード（行）数を確認する場合にも、ソート（並び替え）を元に戻す場合においても、便利です。

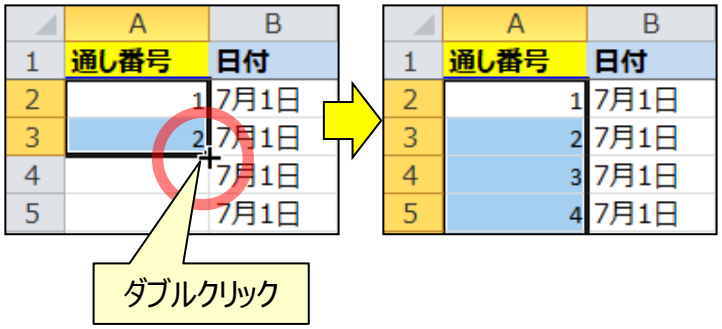
データセットの左端に空白の列を作り、1行目に「1」、2行目に「2」を入力して、入力した二つのセルを選択した状態で、「2」の右下の黒い四角をダブルクリックすると、最終行まで通し番号がつきます。

- 「1」のセルだけを選択した状態で、右下の黒い四角をダブルクリックしてしまうと、最終行まで「1」が並びますので、「2」まで含めて選択してからダブルクリックしてください。

通し番号を挿入したデータセット（10列目まで）

A	B	C	D	E	F
通し番号	日付	曜日	単価	数量	数量
1	7月1日	水	ボールペン黒	100	1
2	7月1日	水	鉛筆	80	5
3	7月1日	水	ボールペン赤	100	1
4	7月1日	水	ボールペン赤	100	2
5	7月1日	水	ノート	150	2
6	7月1日	水	はさみ	400	1
7	7月1日	水	はさみ	400	2
8	7月1日	水	はさみ	400	2
9	7月1日	水	はさみ	400	2
10	7月1日	水	ボールペン赤	100	1

Excelにおける通し番号のつけ方



欠損値・異常値のチェック

◆Excelにおいて、数値の欠損値・異常値の確認をする際は、「ステータスバー」が便利です。

- Excel画面下側の表示倍率を表す枠の部分を**ステータスバー**といい、ステータスバーを右クリックすることで、その表示内容を選択でき、欠損値・異常値の確認に利用できます。
  - データレコードにおいて一部の情報が利用できないものを欠損値または欠測値といいます。欠損値はセルが空白となるケースもあれば、「N/A」と文字列が記入されるケース、数値以外の文字列が入力できなかった場合では「-1」「9999」などの異常値が記入されるケースがあります。

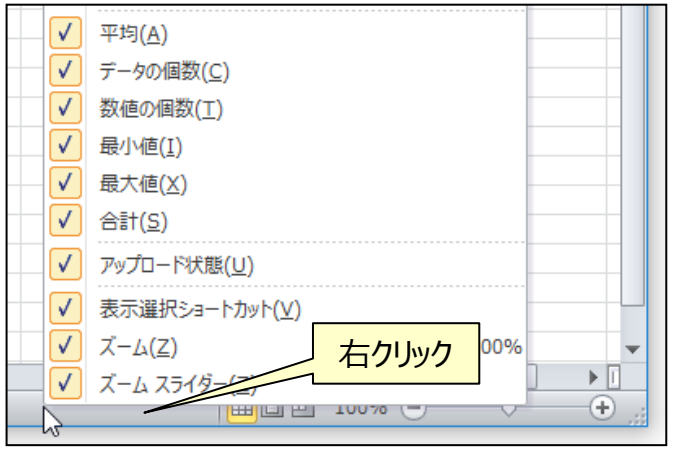
ステータスバーを右クリックし、表示項目の〈平均〉〈データの個数〉〈数値の個数〉〈最大値〉〈最小値〉〈合計〉にチェックを入れます。

- 異常値のチェックに利用するのは、主に〈データの個数〉〈数値の個数〉〈最大値〉〈最小値〉ですが、〈平均〉〈合計〉の表示もデータの確認に便利です。

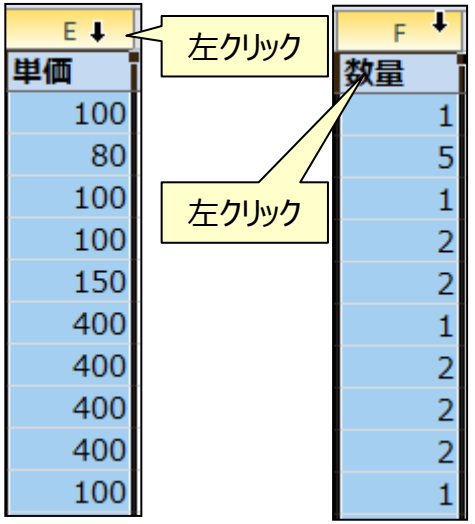
Excelの列頭のアルファベットを左クリックすることで、列全体を選択してからステータスバーを確認します。

- 例示の表のように1行目に［単価］［数量］などの変数名が入っている場合は、〈データの個数〉よりも〈数値の個数〉が1小さくなりますが、それ以上の差があれば、数値が入るべき列に文字入力があり、欠損値の可能性に気が付くことができます。
- ステータスバーの最大値や最小値が現実的な値になっているかを確認することで、簡潔な異常値のチェックができます。

ステータスバーの表示内容を選択



列全体を選択して異常値を確認



E列［単価］選択時のステータスバー

データの個数: 1032　数値の個数: 1000　最小値: 80　最大値: 400

〈データの個数〉よりも〈数値の個数〉が32小さく、
変数名以外に文字の入力が32あることに気がつきます。
最小値、最大値は現実的な値で問題はなさそうです。

F列［数量］選択時のステータスバー

データの個数: 1032　数値の個数: 999　最小値: -1　最大値: 9999

〈データの個数〉よりも〈数値の個数〉が3小さく、
変数名以外に文字の入力が33あることに気がつきます。
最小値、最大値は異常値であることにも気がつきます。

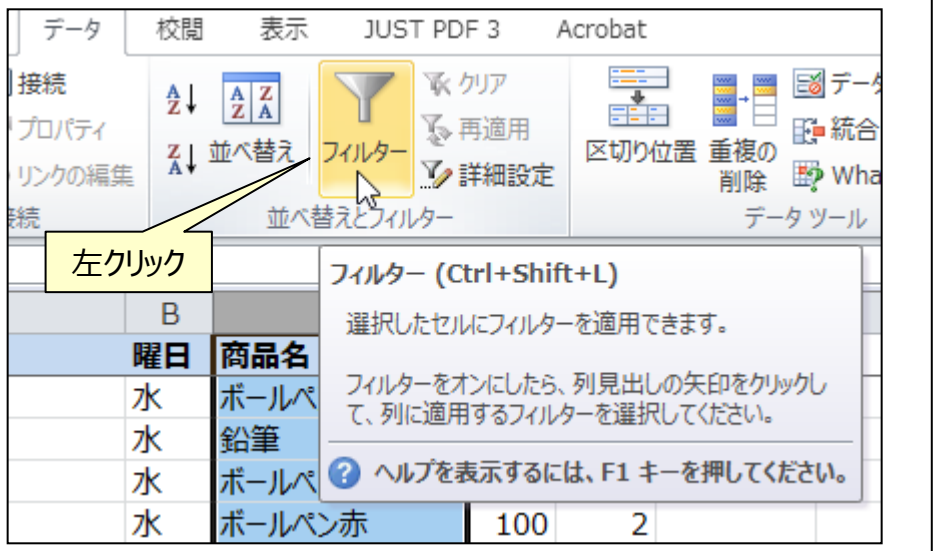
文字列の表記揺れのチェック

◆Excelの「フィルター」を使うと、文字列の表記揺れのパターンを簡単に確認できます。

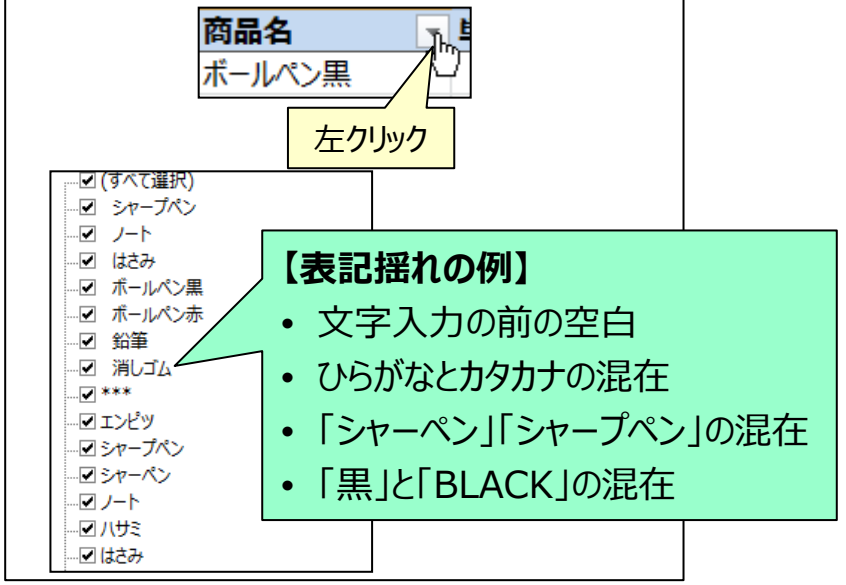
- ［商品名］の列には、「はさみ」と「ハサミ」が混在、文字のはじめに空白が入っているなどの表記揺れがあります。
  - 表記揺れはデータ集計・データ分析の障害となるため、文字列を統一する必要があります。
  - 「フィルター」の本来の用途は、指定した名称に一致する行を選択して抜き出すことですが、名称をリスト化してくれるために表記揺れの確認に利用できます。
- ひらがなとカタカナの混在、空白の挿入、半角全角の相異といった表記揺れの確認には**フィルター**が便利です。
  - Excelのフィルターは、半角と全角は区別する一方で、大文字と小文字は区別せず、大文字小文字の表記揺れはフィルターでは確認できません。

対象とする列の一部のセルを選択した状態で、Excel上部の「データ」タブにある「フィルター」を左クリックして、表示されたメニューからフィルターを左クリックしてください。表記揺れを確認したい列にある（下向きの三角▼）が入った四角をクリックすることで、文字列のリストが表示されます。

対象列を選択した状態で［データ］タブのフィルターをクリック



列頭の箱のボタンをクリック



□ 表記揺れの修正はプロセスが長いため、まずは異常値の確認と修正を行った後に説明します。

欠損値、異常値の置き換え

◆特定の条件に基づくセルの値の変換には、ExcelのIF関数が便利です。

- ExcelのIF関数は『=IF(条件式,条件を満たす場合の出力,条件を満たさない場合の出力)』とコンマで区分して入力することで、条件式で場合分けした出力ができます。
  - 括弧内に対象となる数値やセルを指定することで、定められた処理をするものを関数といいます。Excelではセルに「=」に続いて関数名を記入します。

IF関数で数値を転記し、数値でなければピリオド『NaN』を出力する場合

セルに『=IF(E2<10000,E2," NaN")』と入力すれば、対象セルのE2が考えられる上限の10000より小さい数値であればE2の値をそのまま出力し、文字を含め、それ以外なら『NaN』を出力することで、数値のみを転記できます。

- Excelの条件式において、「記号やスペースを含む全ての文字」はあらゆる数値より大きい値（無限大）として扱われます。このため、文字入力の可能性がある列においては、考えられる下限の0より大きいかを条件とする『=IF(E2>0,E2,"NaN")』ではなく、考えられる上限値（例えば10000）より小さいかを条件とする『=IF(E2<10000,E2,"NaN")』としてください。

数値のみの転記

D	E	F	G
商品名	単価	数量	改訂単価
ボールペン黒	100	1	100
鉛筆	80	5	80
ボールペン赤	100	1	100
ボールペン赤	100	2	100
ノート	150	2	150

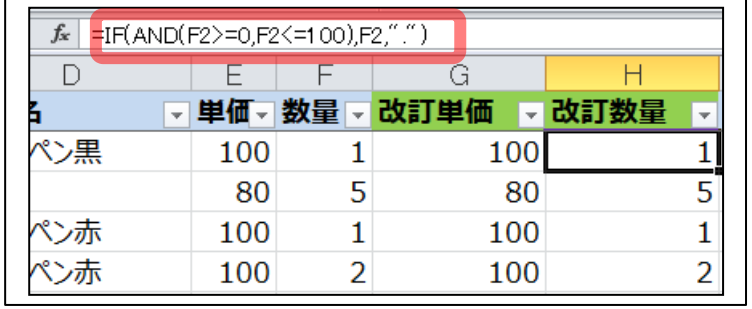
- 欠損値は、プログラミング言語や分析ソフトウェアでの利用も考慮して、『NA』『NULL』『NAN』『.』で表しますが、この講座では『NaN』に置き換えます。

IF関数で-1や9999といった異常値も除き、0以上100以下のみ数値を出力する場合

セルに『=IF(AND(F2>=0,F2<=100),F2,"NaN")』と入力すれば、対象セルのF2が0以上100以下ならE2の値をそのまま出力し、そうでなければ『NaN』を出力することで、数値のみを転記できます。

- 条件式の中に入っているANDは、両方満たす場合の「かつ」を表すExcel関数で、コンマで区切ることで複数の条件を与えることができます。また、どちらかを満たす場合の「または」を表すORというExcel関数もあります。
- Excelの条件式では「より大きい(>)」「より小さい(<)」の記号の後ろにイコール(=)を入れることで、「以上(>=)」「以下(<=)」となります。

0以上100以下の数値の転記



D	E	F	G	H
名	単価	数量	改訂単価	改訂数量
ペン黒	100	1	100	1
	80	5	80	5
ペン赤	100	1	100	1
ペン赤	100	2	100	2

関数を入力後、そのセルの右下の黒い■をダブルクリックすると、下側の列にも同じように関数が入ります。

表記揺れの統一（1）PHONETIC関数の利用

◆PHONETIC関数は、ひらがな、カタカナの表記揺れ統一に利用できます。

- 元の商品名を右側が空白の列にコピーしてから、**１列ずつ右に変換していく形で表記揺れを補正**していきます。
  - 2行目でExcel関数を作った後は、セルの右下の黒い四角■をダブルクリックして、列の最下段まで同じ関数を反映させます。

PHONETIC（フォネティック）関数：文字列の読み仮名をカタカナで出力

セルに『=PHONETIC(I2)』と入力すれば、対象セルI2のフリガナを出力します。

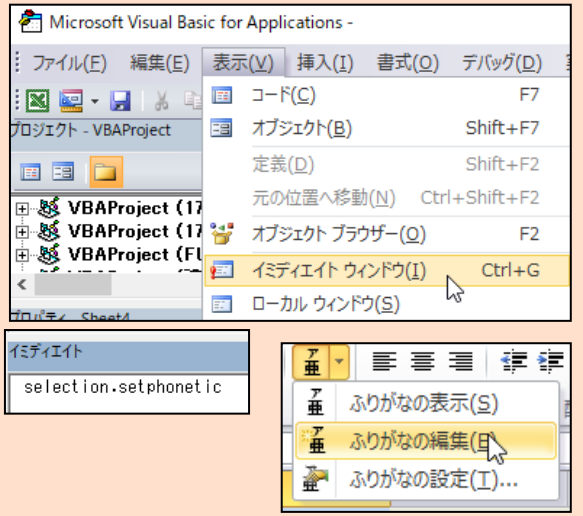
- 初期設定において、PHONETIC関数はひらがな、半角カタカナを全角カタカナで出力します。
- PHONETIC関数は、例外的に関数の出力を引き継げません。例えば、A1に「黒」や「クロ」と入力し、B1に『=PHONETIC(A1)』と入力すれば、「クロ」と出力しますが、C1に『=PHONETIC(B1)』と入力してもB1の「クロ」の出力を引き継がず、C1は空白となります。このため、PHONETIC関数は変換プロセスの最初など、値そのものを括弧内に指定する必要があります。

読み仮名（カタカナ）で統一

【PHONETIC関数における漢字の読み仮名】

◆ PHONETIC関数は、ひらがなとカタカナの統一には常に利用できますが、漢字の読み仮名については、Excelに実際に漢字変換して入力した設定が反映され、他のファイルからコピー&ペーストしたり、CSVを読み込んだ場合は漢字の読み仮名がつかず、漢字のまま出力されてしまいます。

◆ CSV等から読み込んだ漢字に一括して、標準的な読み仮名をつける場合は、読み仮名を付けたい範囲を選択し、「Alt +F11」でVisual Basicを表示し、表示のボタンから「イミディエイト ウィンドウ」を選択し、表示された欄に『selection.setphonetic』と入力し、Enterを押してください。



- フリガナの誤り等は、Excelのメニューの「ふりがなの編集」から変更できます。「ふりがなの設定」ではひらがな表示への変更も可能です。

表記揺れの統一（2）TRIM･UPPER関数の利用

◆TRIM関数は空白の除去、UPPER関数は大文字への統一に利用できます。

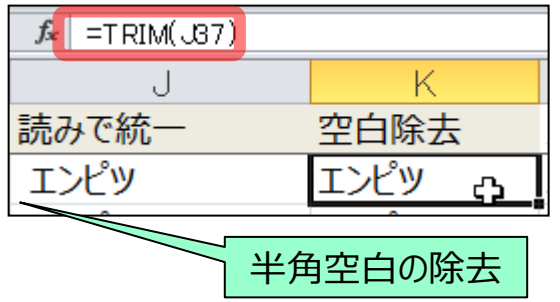
- **TRIM関数**は、文字列の前と後にある全角および半角の空白を除去して出力します。

TRIM（トリム）関数：文字列の始めと終わりの空白を削除して出力

セルに『=TRIM(J2)』と入力すれば、対象セルJ2の前後の空白を除去します。

- TRIM関数は全角の空白、半角の空白をともに除去します。
- TRIM関数は単語内で複数の空白が続く場合は、一つの空白にまとめるため、文字内に空白がある場合は、空白が全てなくなるわけではありません。
  （例）「 ノ　　ート 」→「ノ ート」

空白除去による表記揺れの統一



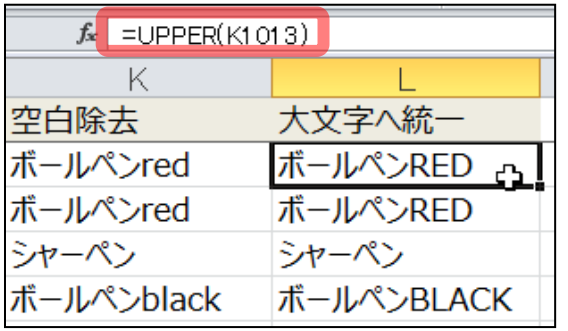
- **UPPER関数**は、英字の小文字を大文字に統一して出力します。

UPPER（アッパー）関数：文字を全て大文字に変更して出力

セルに『=UPPER(K2)』と入力すれば、対象セルK2のアルファベットを全て大文字で統一します。

- Excelの集計において、一般に全角と半角は区別する一方で、大文字と小文字は区別しません。しかし、視覚的な統一感、他のプログラムでの利用可能性を考えれば、大文字と小文字は統一している方が良いです。
- UPPER関数の代わりにLOWER関数を利用すれば、アルファベットを小文字で統一することができます。

アルファベットの大文字への統一



- 半角と全角の混在がある場合は、**ASC関数**で半角に統一するか、**JIS関数**で全角に統一して下さい。
  - 今回のデータクレンジングでは、データチェック時に半角・全角の不統一がなかったため、ASC関数の利用は省略しています。

表記揺れの統一（3）SUBSTITUTE関数の利用

◆SUBSTITUTE関数は、文字の置き換えに利用できます。

- ExcelのSUBSTITUTE関数は『=SUBSTITUTE(対象となるセルの番地,“置き換え元の文字列”,“置き換え後の文字列”)』と、コンマで区切り、引用符で文字列を区切って指定します。

SUBSTITUTE（サブスティチュート）関数：文字を置き換えて出力

セルに『=SUBSTITUTE(N2,“シャーペン”,“シャープペン”)』と入力すれば、対象セルN2の「シャーペン」という文字列を「シャープペン」に置き換えます。

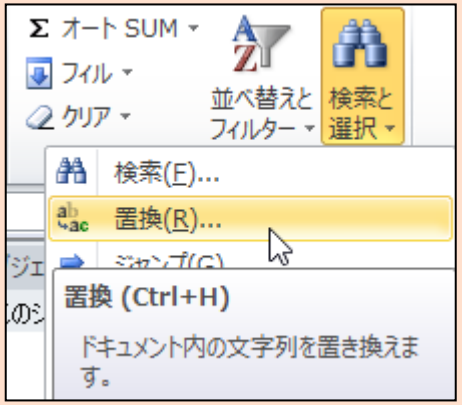
- 削除したい文字列がある場合は『=SUBSTITUTE(N2,“[削除対象文字列]”,“”)』とすることで、文字列を削除できます。

置き換えによる表記揺れの統一

【Excelメニューの検索・置換機能を利用したデータクレンジングについて】

- ◆ Excelの「検索と置換」から置換機能を使うことでも、表記揺れの統一作業は可能ですが、Excel関数を使う場合に比べて、作業手順や置換内容が分かりにくくなります。
- ◆ 本講座のようにExcel関数を使って置換すると、作業手順、置換内容を簡単に確認することができ、事後的に追加するデータレコードがある場合やデータクレンジングの方法を変更したい場合でも対応が容易です。

それでもExcelの置換機能を使う場合は、置換した文字のリスト、指定範囲を別シート等に記録するようにしましょう。



クレンジング完了の確認と値での貼り付け

◆データクレンジング後（改訂後）の列を関数との関係が切れた「値」で貼り付けます。

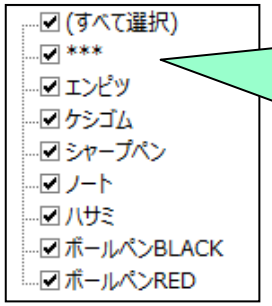
- 異常値や表記揺れの改訂が完了したことを、ステータスバーやフィルターから確認します。

クレンジング前のF列［数量］の選択時のステータスバー

クレンジング後のH列［改訂数量］の選択時のステータスバー

［改訂数量］は、-1や9999といった［数量］の異常な値を一つずつ、文字列の『NaN』に変更し「数値の個数」は2減少し、最大値、最小値も正常な範囲にあることを確認できます。

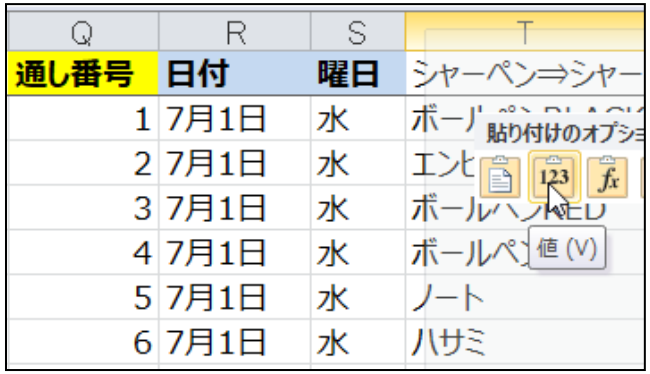
O列より、商品名リストの表記統一を確認



日付の区切りの「***」と各商品の統一された名称のみが表示されており、表記揺れが解消していることを確認できます。

- 分析等に利用する列の貼り付け先を決め、元のデータセットと1列以上空けるか別シートに**「値」で貼り付け**ます。
  - 「値での貼り付け」は、コピーしてから右クリックメニューで指定します。値で貼り付けると、計算や変換に利用したExcel関数との関係が切れるため、改めてExcel上のデータ分析が可能になるとともに、他の分析用のプログラム言語、ソフトウェアでも利用できます。
  - セルの色や表示形式も貼り付けたい場合は、いったん「Ctrl+V」の通常の貼り付けで書式等を含めて貼り付けた範囲に、改めて「値」で貼り付けます。
  - Excel上で1列をあけると、視覚的にも区切りが明らかになることに加えて、フィルター等でも別のデータセットとして認識されます。

クレンジング完了の列を値で貼り付け



Q	R	S	T
通し番号	日付	曜日	シャーペン⇒シャー
1	7月1日	水	ボール
2	7月1日	水	エン
3	7月1日	水	ボール
4	7月1日	水	ボール
5	7月1日	水	ノート
6	7月1日	水	ハサミ

表記揺れと異常値を除いた改訂済データセット（1列をあけて貼り付け）

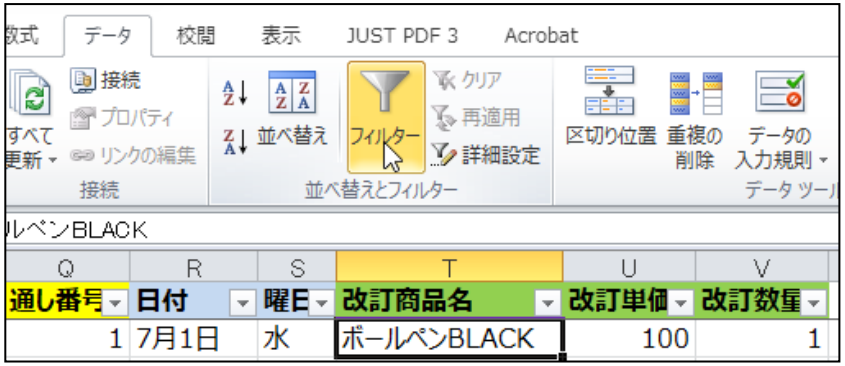
P	Q	R	S	T	U	V
	通し番号	日付	曜日	改訂商品名	改訂単価	改訂数量
	1	7月1日	水	ボールペンBLACK	100	1
	2	7月1日	水	エンピツ	80	5
	3	7月1日	水	ボールペンRED	100	1
	4	7月1日	水	ボールペンRED	100	2
	5	7月1日	水	ノート	150	2
	6	7月1日	水	ハサミ	400	1

フィルターによる不要データレコード（行）の非表示

◆Excelのフィルター機能を使って、分析に利用するデータレコード（行）のみを表示します。

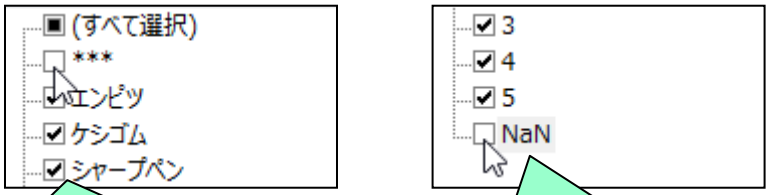
- **フィルター本来の用途**として、区切り文字や欠損値のリストのチェックマークを外し、**表示対象から除外**します。
  - 本講座では、これまでに「表記揺れの確認」のためにフィルターの表示を利用しました。
  - 異常値または欠損値があり、正しい数値が分からない場合の対処方法は、大別して二種類あります。一つの対処方法は、利用が困難なデータレコードを分析対象とするデータセットから除去すること、もう一つの対処方法はもっともらしい値を入力することです。この講座では、より簡単な対処方法として、異常値・欠損値のデータレコードを分析対象とするデータセットから除去する対処方法を示します。

改訂データセットに対してフィルターを利用



Q	R	S	T	U	V
通し番号	日付	曜日	改訂商品名	改訂単価	改訂数量
1	7月1日	水	ボールペンBLACK	100	1

区切りの「***」と異常値・欠損値の「NaN」を除外して表示



［改訂商品名］のフィルターから、日付の区切りに相当する「***」の左側のチェックマークを外します。

［改訂数量］のフィルターから、異常値を変換した『NaN』の左側のチェックマークを外します。

- データセットにおける区切り文字や欠損値のデータレコードが非表示となっていることを確認します。
  - フィルターによって非表示の行（データレコード）がある場合は、Excelの左側の行番号を示す文字が青色になります。
  - Excelの左側の行表示やの［通し番号］の値が飛んでいることで、オリジナルのデータセットから非表示になっている行があることが分かります。

フィルターをかける前（全ての列の表示）

	Q	R	S	T	U	V
1	通し番号	日付	曜日	改訂商品名	改訂単価	改訂数量
34	33	7月1日	水	ボールペンBLACK	100	1
35	34	7月1日	水	ハサミ	400	NaN
36	35	***	***	***	NaN	NaN
37	36	7月2日	木	エンピツ	80	1
38	37	7月2日	木	エンピツ	80	1

フィルターをかけた後（区切り行、欠損値の非表示）

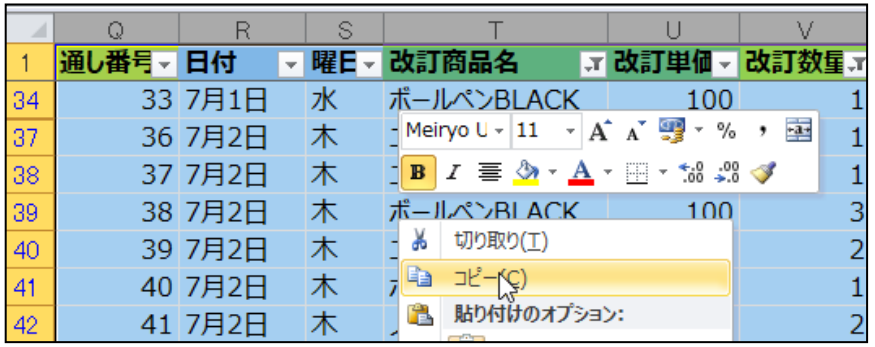
	Q	R	S	T	U	V
1	通し番号	日付	曜日	改訂商品名	改訂単価	改訂数量
34	33	7月1日	水	ボールペンBLACK	100	1
37	36	7月2日	木	エンピツ	80	1
38	37	7月2日	木	エンピツ	80	1
39	38	7月2日	木	ボールペンBLACK	100	3
40	39	7月2日	木	エンピツ	80	2

分析に利用するデータレコード（行）の抽出

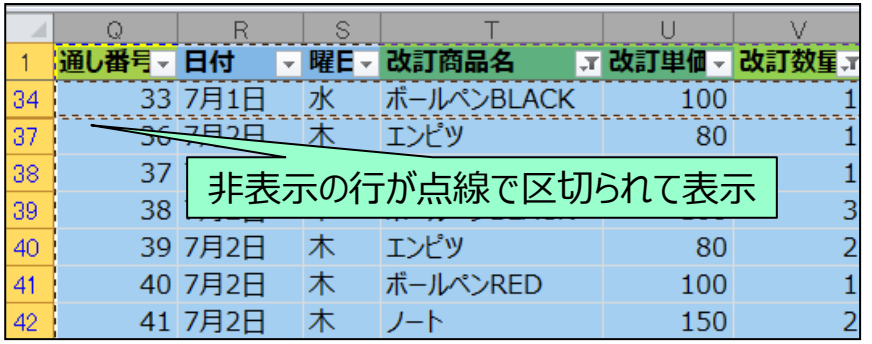
◆不要な行を除去して、分析に使うデータレコードだけを別シートに貼り付けます。

- フィルターによって分析対象とするデータレコード（行）のみを表示した状態で、それらの列を全て選択し「コピー」をクリックすると、**非表示の行は点線で区切られてコピー範囲が表示**されます。

フィルターをかけた状態で「コピー」を選択



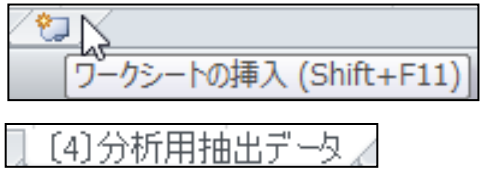
非表示の行をコピー対象に含めない点線での表示



	Q	R	S	T	U	V
1	通し番号	日付	曜日	改訂商品名	改訂単価	改訂数量
34	33	7月1日	水	ボールペンBLACK	100	1
37	36	7月2日	木	エンピツ	80	1
38	37	7月2日	木			1
39	38	7月2日	木			3
40	39	7月2日	木	エンピツ	80	2
41	40	7月2日	木	ボールペンRED	100	1
42	41	7月2日	木	ノート	150	2

- Excelのシート群の右側にあるボタンをクリックして新規作成したシートに、コピーしたデータレコードを貼り付けます。
  - シート内で行（データレコード）の一貫性を確保することに加えて、データクレンジング作業と分析作業を区分するために別シートに貼り付けます。
  - クレンジング前データと抽出データの行の対応、除外された行を分かりやすくするためにも、講座の冒頭部で［通し番号］をつけていました。

新規作成したシートへの必要なデータレコードのみを貼り付け



	A	B	C	D	E	F
1	通し番号	日付	曜日	改訂商品名	改訂単価	改訂数量
34	33	7月1日	水	ボールペンBLACK	100	1
35	36	7月2日	木	エンピツ	80	1
36	37	7月2日	木			1
37	38	7月2日	木			3
38	39	7月2日	木	エンピツ	80	2
39	40	7月2日	木	ボールペンRED	100	1

分析対象から外した行（データレコード）を除外して貼り付けることができます。

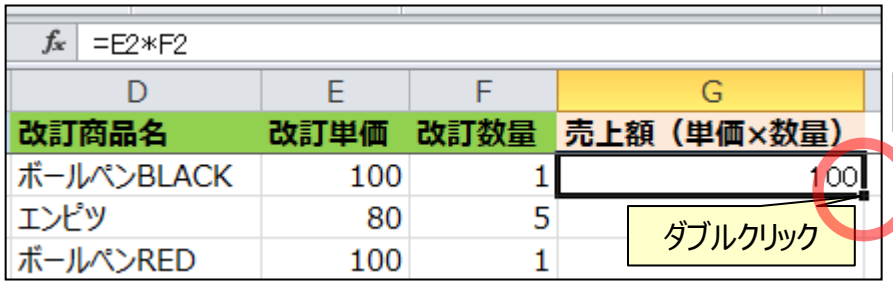
- 以上でデータクレンジングと分析用データレコードの抽出を終え、以降ではデータ集計を行います。

全体の合計値の導出：SUM関数

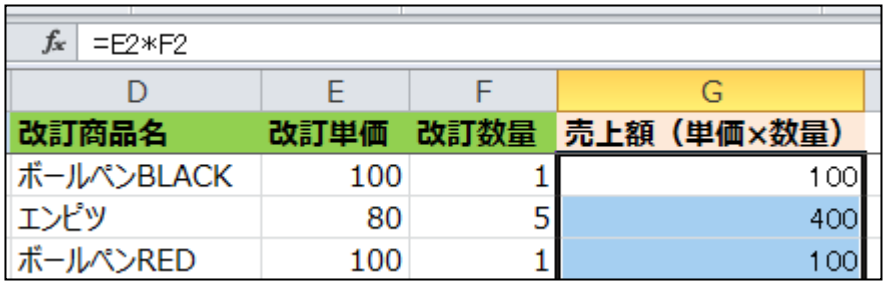
◆売上額の列に対してSUM関数を用いて、合計値を導出します。

- 個々の行にある単価と数量を掛け合わせて［売上額］を作成します。G2セルに『=E2*F2』と入力します。
  - データクレンジングの完了後はExcel上のピボットテーブルを使うケースもあれば、統計分析用のプログラミング・ソフトウェアで処理することもあります。Excelのピボットテーブルを利用したデータ処理は講座3-3、プログラミングソフトRを使ったデータ処理は講座4-3で示します。
- G2セル選択時に右下に表示される■をダブルクリックすることで、G列の下部に同じルールの計算を適用します。

G2セルにおける売上額（単価×数量）の導出



右下の■をダブルクリックすることで全行に適用



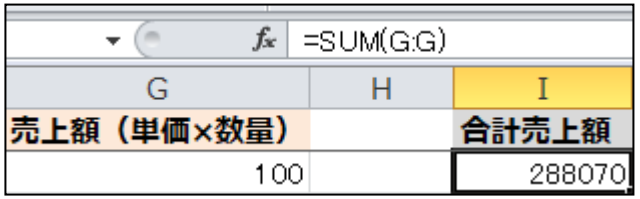
- 売上の列全体を範囲指定した際のステータスバーの合計値および合計値を導出する**SUM関数**によって、［合計売上額］を把握できます。
  - SUM関数では文字列は除外して計算されるため、変数名を含んで範囲指定をすることができます。Excelでは(左端の列:右端の列)の形で範囲指定ができるので、G列全体の合計値を導出する場合は、『=SUM(G:G)』と指定してください。

SUMIF（サムイフ）関数：対象範囲の合計値を算出

ステータスバーにおける売上高に関する表示

=

SUM関数での合計値の出力



□ 続いて、日付別や商品別に売上を集計するために、各項目のリストを作成します。

「重複の削除」による項目名リスト作成

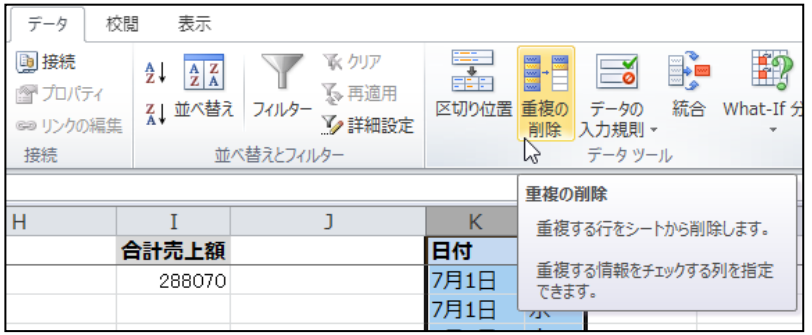
◆条件別の集計値を導出するために、「重複の削除」によって、項目名リストを作成します。

- Excel関数で条件別の集計値を導出するためには、集計対象とする項目名をリストアップする必要があります。
  - 条件別の集計値は講座3-3で示すピボットテーブルでも導出できますが、ここではExcel関数を利用した条件別の集計値の導出を示します。
- 項目名リストを作成するためにB列とC列の［日付］と［曜日］をK列とL列へ、2列あけてD列の［改訂商品名］をO列に貼り付け、各列を選択した状態でExcelのメニューの「データ」タブにある**「重複の削除」**をクリックします。

項目名リスト用の列の配置

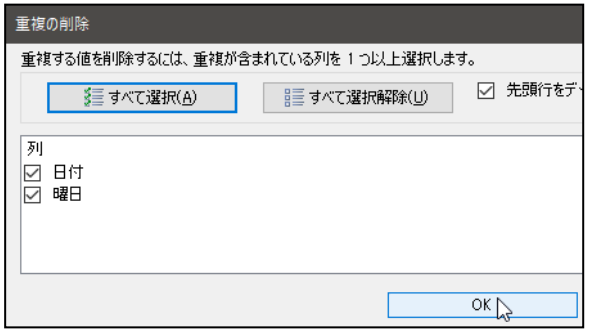
K	L	M	N	O
日付	**曜日**			**改訂商品名**
7月1日	水			ボールペンBLACK
7月1日	水			エンピツ
7月1日	水			ボールペンRED
7月1日	水			ボールペンRED
7月1日	水			ノート

K列とL列を選択して「重複の削除」をクリック



- 表示されたダイアログボックスの設定を変えずに「OK」をクリックすると、それぞれの項目名リストが表示されます。
  - 「重複の削除」は、本来、指定した変数の組み合わせから重複データを除外するための機能ですが、項目名リスト作成にも利用できます。
  - 「データレコードのIDが一致している」などの重複記入があるデータセットの場合、「重複の削除」はデータクレンジングにおいても利用します。

ダイアログボックスで「OK」をクリック



重複削除に関する情報表示

「重複の削除」は記録が残らないため、
操作内容を記録しておくことが望ましいです。

［日付と曜日］と［商品名］に関して、
重複と漏れのない項目名リストが完成

［日付］と［曜日］と［改訂商品名］のリスト

K	L	M	N	O
日付	**曜日**			**改訂商品名**
7月1日	水			ボールペンBLACK
7月2日	木			エンピツ
7月3日	金			ボールペンRED
7月4日	土			ノート
7月5日	日			ハサミ
7月6日	月			ケシゴム
7月7日	火			シャープペン
7月8日	水			

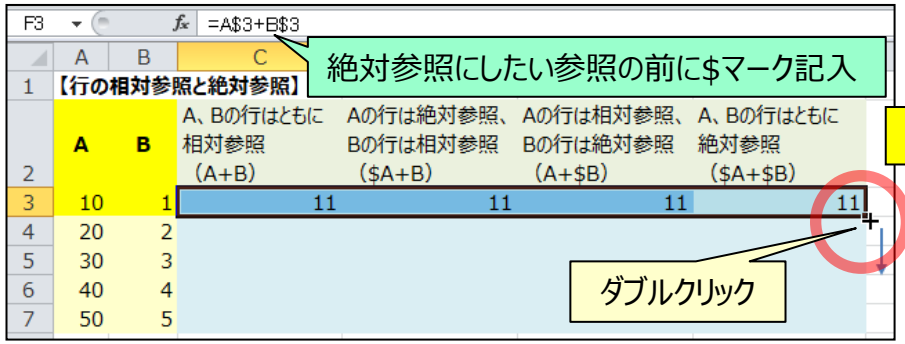
【参考】Excelにおける相対参照と絶対参照

◆Excelにおけるセルの参照には、相対参照と絶対参照の2種類があります。

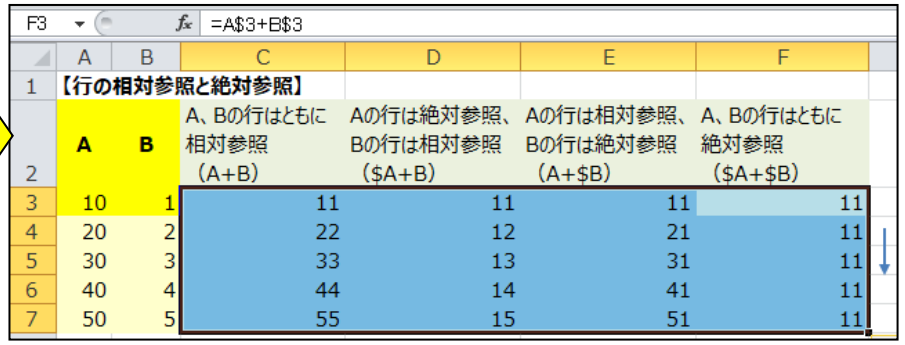
- 次のVLOOKUP関数の照合対象範囲の設定の仕方に関連して、Excelの**相対参照**と**絶対参照**を紹介します。
- Excelには、セルのコピーや入力の引き延ばしに応じて、**演算や関数が参照するセルが対応する形で移動する相対参照**と**参照するセルが固定されている絶対参照**があります。
  - Excelの標準設定は相対参照であり、絶対参照を指定する操作をしなければ、自動的に相対参照になっています。
  - 絶対参照を行うには、演算やExcel関数で他のセルを参照しているセルの記入に$（ドル記号）を挿入します。
- 実習用Excelのシート〔参考〕では、「行」「列」「行・列」の3種で、相対参照と絶対参照の違いを確認できます。

シート〔参考〕のC3～F3セルを選択した状態で、F3の右下に表示される■をクリックして確認してください。

シート〔参考〕の行の相対参照と絶対参照



相対参照と絶対参照の組み合わせによる出力



シート〔参考〕のJ列およびP列の青いセルの範囲を選択して、水色の範囲に入力を引き延ばして、出力を確認してください。

- 演算や関数で参照するセルの指定の前に**「$（ドル記号）」を記入すると、絶対参照**になり、セルのコピーや入力の引き延ばしをしても、参照するセルが変わらず固定されています。
  - 横の行のみ絶対参照にする場合は「A$1」、縦の列のみ絶対参照にする場合は「$A1」という形式で指定し、行と列の双方を固定する場合は「A1」と固定したい行または列の前に$を記入します。

データセットの結合（1）VLOOKUP関数の利用

◆VLOOKUP関数を使って、データセットの結合や変数の追加ができます。

- シート〔4〕と〔5〕のAA・AB列には日付別天気データがあり、日付・曜日のリストの右側に天気のデータを結合します。
- 天気データを日付データに結合するためには、Excelの**VLOOKUP関数**を利用します。
  - 今回の実習用データでは、K列は重複のない日付順になっているのでAB列の該当部分を貼り付けることでもデータセットを結合できますが、VLOOKUP関数は、照合する文字に重複があったり、照合する文字が順不同になっている項目であってもデータセットを結合できます。
- ExcelのVLOOKUP関数は『=VLOOKUP(照合する文字,照合対象範囲,照合対象範囲内の表示する列目,近似値の可否)』と、コンマで区切って指定することで、「照合する文字」が「照合対象範囲」の一列目が一致した場合に、「照合対象範囲内の表示する列目」の文字列を表示します。

M2のセルに『=VLOOKUP(K2,AA2:AB36,2,FALSE)』と入力すると、7月1日の天気データとして「晴れ」と表示されます。

- Excel関数における範囲の指定は、「左上のセルの番地：右下のセルの番地」として指定します。「AA2:AB36」という指定で、6月29日～8月2日の天気データが記載された35行2列の長方形の範囲指定となります。なお「AA:AB」とセルの番地の数値の記入を抜けば、列全体の指定になります。
- VLOOKUP関数の第1項目の「K2」にある文字「7月1日」と第2項目の「AA2:AB43」の範囲の1列目の「6月29日～8月2日」の文字列を照合し、第4項目によって近似値を許可せずに完全一致の文字列がある行において、第3項目の「2」列目の文字をセルに出力します。
- 近似値の可否は、許可する場合は「TRUE」、許可しない場合は「FALSE」と入力します。データセットの結合では、原則として「FALSE」にしてください。

VLOOKUP関数を使用して対応する「天気」を出力

M2 =VLOOKUP(K2,AA2:AB43,2,FALSE)

	J	K	L	M	N
1		日付	曜日		
2		7月1日	水	晴れ	
3		7月2日	木		
4		7月3日	金		
5		7月4日	土		

照合する文字の列　関数の出力の列

照合対象範囲の［日付］と［天気］のリスト

AB4 晴れ

	Z	AA	AB	AC	AD
1		日付	天気		
2		6月29日	くもり		
3		6月30日	くもり		
4		7月1日	晴れ		
5		7月2日	くもり		

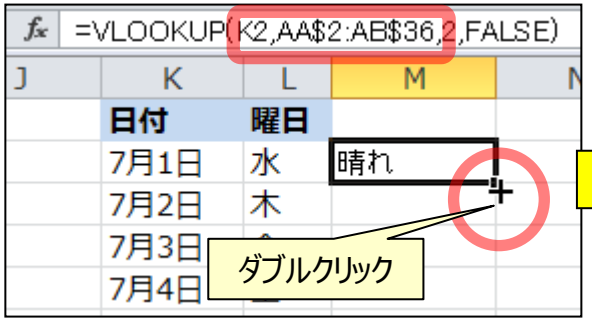
照合対象範囲の1列目　照合対象範囲の2列目

データセットの結合（2）VLOOKUP関数の指定

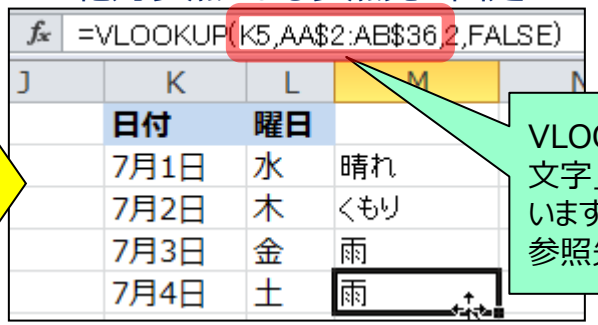
◆VLOOKUP関数の照合対象範囲は、絶対参照での指定が便利です。

- VLOOKUP関数を利用したM2のセルの『=VLOOKUP(K2,AA2:AB36,2,FALSE)』の第2項目の「照合対象範囲」の行を絶対参照に変更すべく『=VLOOKUP(K2,AA**$**2:AB**$36**,2,FALSE)』と$マークを挿入します。
  - 今回の実習用データでは、結合・追加したい変数が天気データの一種類のみであるため、照合対象範囲の列側を絶対参照で固定する必要はありません。もし、結合・追加したい変数が複数ある場合は、第2項目は行・列ともに絶対参照で照合対象範囲を固定して、第3項目の数字を増やすことで複数の変数を結合・追加します。
- M2セルの右下の■をダブルクリックするか、下側へドラッグすることによって、3行目以下のセルも同様に入力します。

絶対参照での入力の引き延ばし



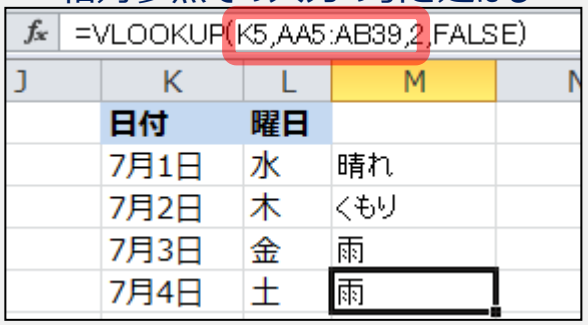
絶対参照による参照先の固定



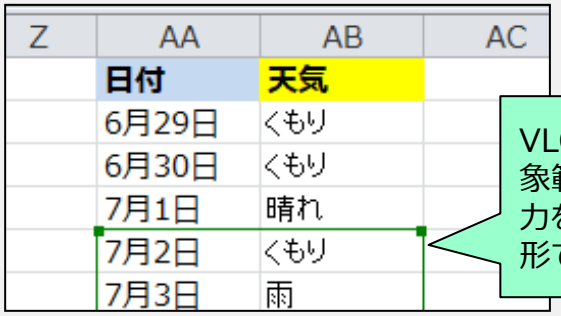
VLOOKUP関数の第1項目の「照合する文字」は、相対参照で参照先が移動していますが、第2項目の「照合対象範囲」の参照先は絶対参照で固定されています。

□ VLOOKUP関数の第2項目を相対参照のままで入力を引き延ばすと、照合対象範囲の参照先がずれてしまいます。

相対参照での入力の引き延ばし



相対参照による参照先のずれ



VLOOKUP関数の、第2項目の「照合対象範囲」が相対参照のままでは、M列の入力を引き延ばした際に、参照先も対応する形で移動して、ずれてしまいます。

項目名別合計値の導出：SUMIF関数

◆SUMIF関数は、項目別の合計値を導出できます。

- ExcelのSUMIF関数は『=SUMIF(照合対象範囲,照合する文字,合計対象範囲)』と、コンマで区切って指定することで、「照合対象範囲」と「照合する文字」が一致した行に関して、「合計対象範囲」の合計値を導出します。
- 日付別の合計売上額を導出するために、「照合対象範囲」をデータセットの日付のB列、「照合する文字」を項目名リストの重複のない日付のK列、「合計対象範囲」をデータセットの売上額のG列に設定します。

SUMIF（サムイフ）関数：「照合対象範囲」と「照合する文字」が一致した行で合計値を算出

「照合対象範囲」および「合計対象範囲」の最上段、最下段の行を絶対参照で指定し、『=SUMIF(B$2:B$998,K2,G$2:G$998)』と入力します。

2行目の日付別合計額が正しく表示されれば、セルの右下の■をダブルクリックするか、下側に延ばして3行目以下も同様に入力します。

- 照合対象範囲や合計対象範囲を列全体とする場合は、絶対参照とは別の指定方法として『=SUMIF(B:B,K2,G:G)』と指定する方法もあります。相対参照による問題は、指定セルの移動によって参照先の行が変わりますが、列全体を指定すると変化する行自体がありません。

［日付別合計額］の表示

=SUMIF(B$2:B$998,K2,G$2:G$998)

K	L	M	N
日付	曜日	結合：天気	日付別合計額
7月1日	水	晴れ	11020
7月2日	木	くもり	10450

代替的な指定方法：=SUMIF(B:B,K2,G:G)

- 同様の手順で項目名リストの商品名の右側に、商品別販売数および商品別合計数を導出します。

商品別売上数は、先頭行に「照合する文字」を「ボールペンBLACK」のセル番地N2とする『=SUMIF(D$2:D$998,N2,F$2:F$998)』と入力し、「ボールペンBLACK」の売上数の表示後にセル右下の■をダブルクリックします。

商品別合計額は、先頭行に「照合する文字」を「ボールペンBLACK」のセル番地N2とする『=SUMIF(D$2:D$998,N2,G$2:G$998)』と入力し、「ボールペンBLACK」の合計額の表示後にセル右下の■をダブルクリックします。

［商品別売上数］［商品別合計額］の表示

=SUMIF(D$2:D$998,O2,F$2:F$998)

O	P	Q
改訂商品名	商品別売上数	商品別合計額
ボールペンBLACK	471	47100
エンピツ	339	27120
ボールペンRED	164	16400
ノート	303	45450

項目名別平均値の導出：AVEREGEIF関数

◆AVERAGEIF関数は、項目別の平均値を導出できます。

- ExcelのAVERAGEIF関数は『=AVEREGEIF(照合対象範囲,照合する文字,平均対象範囲)』と、コンマで区切って指定することで、照合対象範囲と照合文字が一致した行に関して、平均対象範囲の平均値を導出します。
- ［曜日別平均売上額］および［天気別平均売上額］を導出するために、S列に「月、火、水、木、金、土、日」の［曜日］、U列に「晴れ、くもり、雨」の［天気］の項目を文字列で記入し、項目名別リストの合計売上額をAVERAGEIF関数で指定します。

AVERGEIF（アベレージイフ）関数：「照合対象範囲」と「照合する文字」が一致した行で平均値を算出

［曜日別平均売上額］は、先頭行に照合する文字を「月」のセル番地S2とする『=SUMIF(L$2:$L$32,S2,N$2:N$32)』と入力し、「月」の平均売上額の表示後にセル右下の■をダブルクリックします。

［天気別平均売上額］は、先頭行に照合する文字を「晴れ」のセル番地S2とする『=SUMIF(M$2:M$32,U2,N$2:N$32)』と入力し、「晴れ」の平均売上額の表示後にセル右下の■をダブルクリックします。

［曜日別平均売上額］および［天気別平均売上額］の表示

T2　=AVERAGEIF(L$2:L$32,S2,N$2:N$32)

	R	S	T	U	V
1		曜日	曜日別平均売上額	天気	天気別平均売上額
2		月	8418	晴れ	10859
3		火	9303	くもり	8901
4		水	10568	雨	7857
5		木	9994		
6		金	8868		
7		土	9843		
8		日	7668		

［曜日別平均売上額］の出力より、水曜日が最も平均売上額が高く、日曜日が最も平均売上額が低いことが分かります。

［天気別平均売上額］は、晴れの日で10,859円、くもりの日で8,901円、雨の日で7,857円です。

時間の経過に関するデータの可視化：折れ線グラフ

◆時間の経過に関するデータの可視化は、折れ線グラフが適しています。

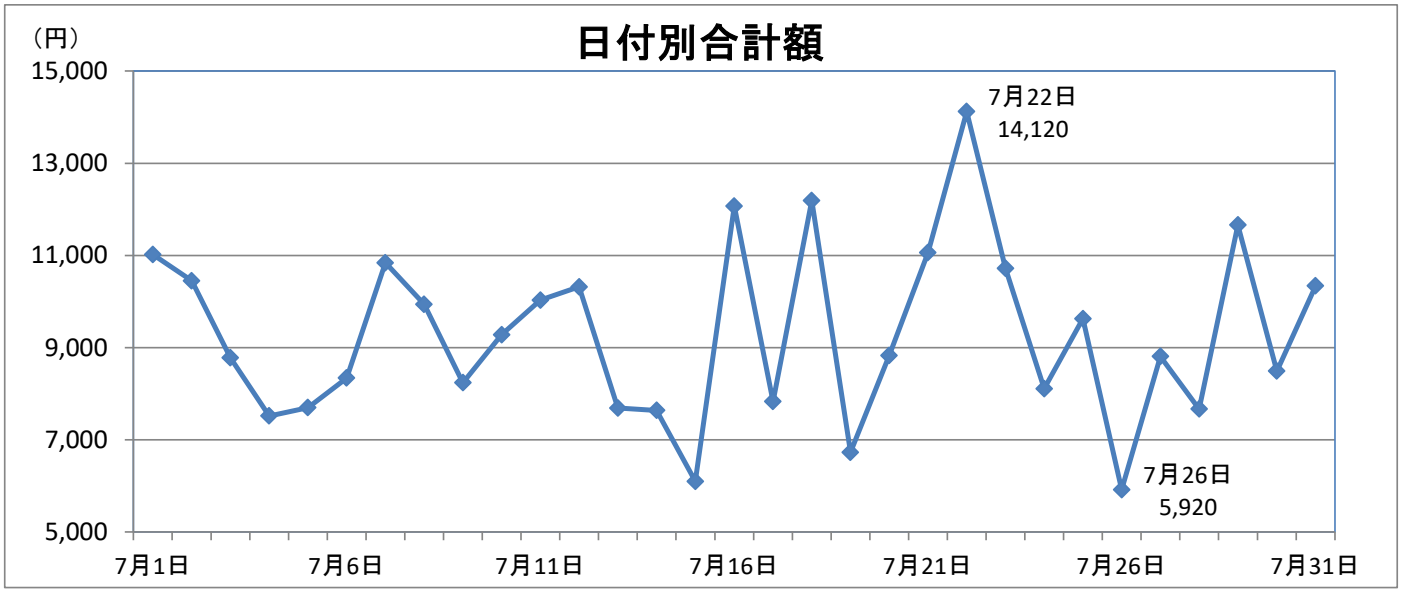
- 推移などの**時間の経過に関するデータの可視化は折れ線グラフ**が適しており、［日付別合計額］を可視化します。
  - 集計表も可視化の一種ですが、画像で把握できるグラフでの可視化の方が理解しやすく、印象に残りやすくなります。

グラフの横軸にしたい日時の情報と数値のデータを縦に並べ、範囲を選択した状態で、Excelの「挿入」タブのグラフ「折れ線」内にある「折れ線」または「マーカー付き折れ線」をクリックします。

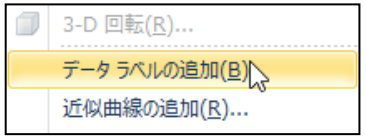
集計表による可視化

日付	日付別合計額
7月1日	11,020
7月2日	10,450
7月3日	8,780
7月4日	7,520
7月5日	7,700
7月6日	8,340
7月7日	10,840
7月8日	9,940
:	:

折れ線グラフによる推移の可視化



- 右上のグラフでは最大値と最小値のデータラベルを折れ線グラフ内に記入しています。
  - グラフ内の各データを表すマーカーを右クリックして「データラベルの追加」をクリックすると、グラフ内にデータの値が表示されます。
  - 全てのマーカーが選択された状態で「データラベルの追加」を選択することで、一括してデータラベルを入れることができます。
  - 表示されたデータラベルを右クリックし、「データラベルの書式設定」を選択することで、データラベルの表示内容や区切り方を選択できます。



割合に関するデータの可視化：円グラフ

◆割合に関するデータの可視化は、円グラフが適しています。

- 割合の**構成表示は円グラフ**が適しており、［商品別合計額］の構成を円グラフで可視化します。
- 円グラフは割合が大きい順に表示すると分かりやすいため、まず［商品別合計額］が大きい項目順に並べ替えます。

並べ替えたい変数を指定して、「データ」タブの「並べ替え」をクリックして、「降順（大きい方から降りていく順）」に並べ替えます。

Excelの機能における「並べ替え」による「商品別合計額」の降順の並び替え

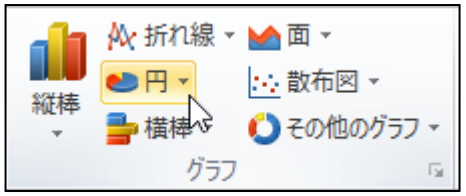
「並べ替え」は記録が残らないため、操作内容を記録しておくことが望ましいです。

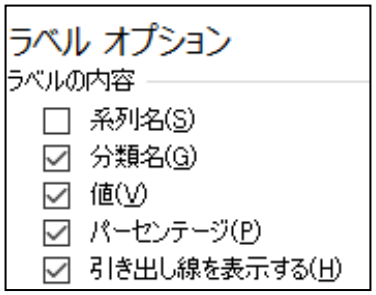
割合を区分する商品名の情報と割合を表したい商品別合計額の情報を縦に並べ、範囲を選択した状態でExcelの「挿入」タブのグラフ「円」をクリックします。

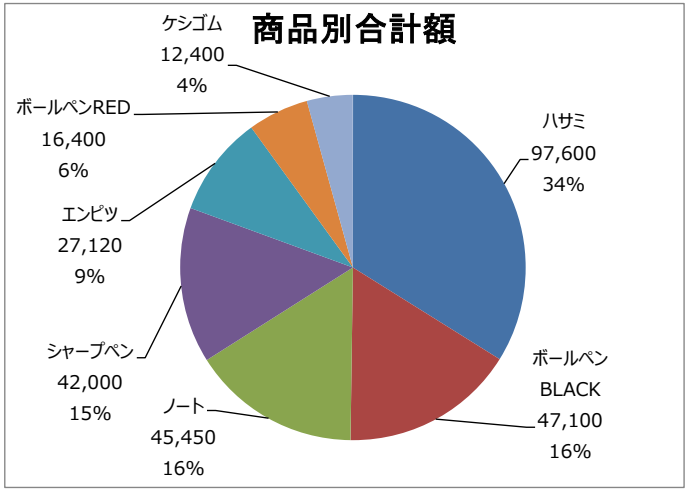
- 円グラフを右クリックして「データラベルの追加」をクリックし、「値」にチェックすると額が表示され、「パーセンテージ」にチェックすると割合が表示されます。
- データラベルの「引出線を表示する」によって、グラフの外側の白地の部分にラベルの情報を記入できます。

「商品別合計額」が大きい順に並べ替えたデータに基づく円グラフの作成

商品名	商品別合計額
ハサミ	97,600
ボールペンBLACK	47,100
ノート	45,450
シャープペン	42,000
エンピツ	27,120
ボールペンRED	16,400
ケシゴム	12,400

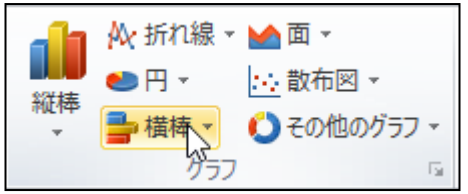






項目間の差を示すデータの可視化：棒グラフ

◆水準や項目間の差に関する表示は、棒グラフが適しています。



- 異なる**項目の水準の差を示す可視化は棒グラフ**が適しています。

  分類としたい項目のテキストと水準を表したい数値を縦に並べ、範囲を選択した状態でExcelの「挿入」タブのグラフ「縦棒」または「横棒」クリックします。

  - ここでは例示として、横棒のグラフを使って［曜日別平均売上額］の可視化を説明します。

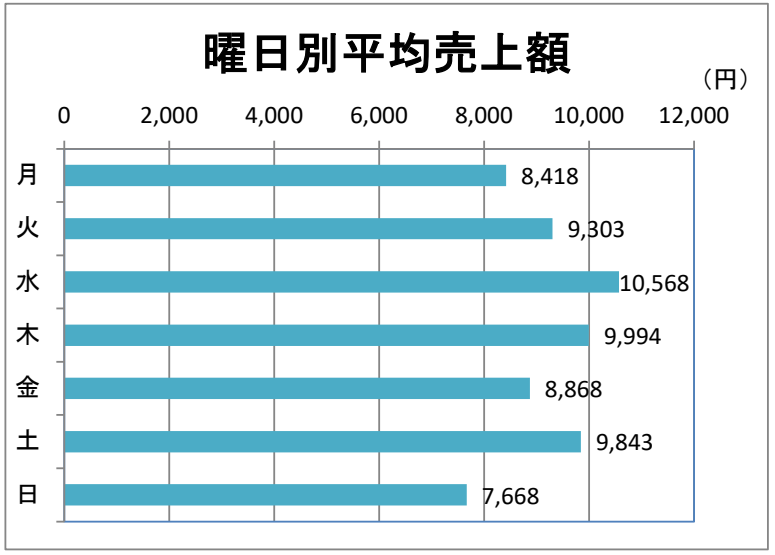
- 可視化においては、作成者が強調したい点を図表に入れ込むことができるケースがあります。
  - 営業資料などの作成者の主観的な強調が許容されるケースもあれば、学術資料などの客観的・画一的な表示が望ましいケースもあります。

- 下記の3種の図表が持っている情報量は同じですが、情報の分かりやすさ、印象は異なります。
  - グラフの最大値、最小値といった表示範囲を変更し、縮尺を変えるだけでも、グラフの印象が変わります。

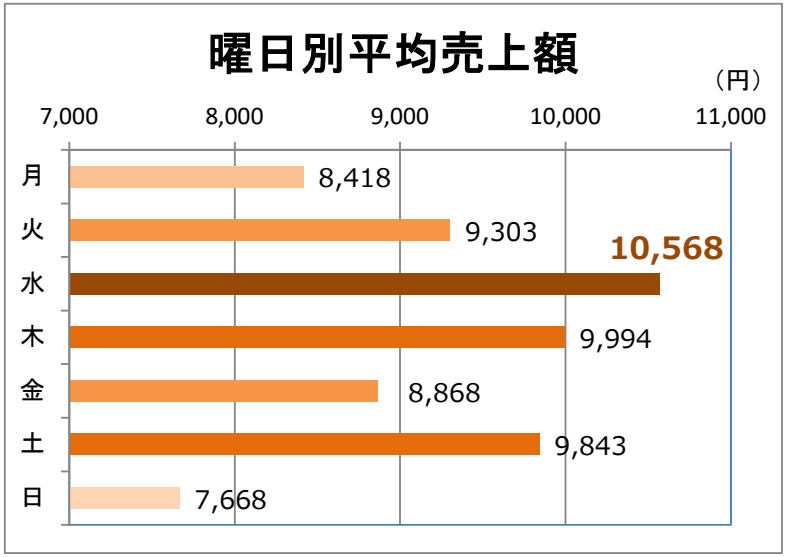
表による可視化

曜日	曜日別 平均売上額
月	8,418
火	9,303
水	10,568
木	9,994
金	8,868
土	9,843
日	7,668

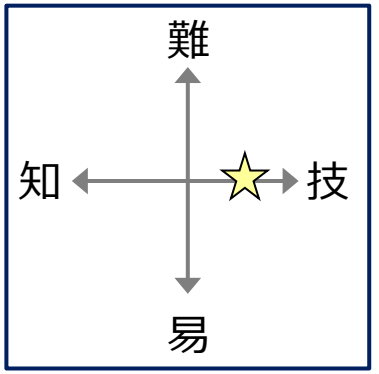
客観的・画一的な棒グラフ



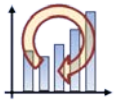
尺度を変更し、強調表示をした棒グラフ



総務省　ICTスキル総合習得教材



［コース3］データ分析



3-3：基本統計量･クロス集計表の作成

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_3_3.pdf

	1	2	3	4	5
［コース1］データ収集					
［コース2］データ蓄積					
［コース3］データ分析			●		
［コース4］データ利活用					

本講座の学習内容［3-3：基本統計量･クロス集計表の作成］

【講座概要】

- 変数の特徴を示す基本統計量を示し、Excel関数における導出方法を紹介します。
- 4種類の数値の尺度と代表的な値の表示に適した基本統計量を示します。
- 代表的な値以外を示す基本統計量として、変数のばらつきや分布を示す指標を紹介します。
- Excelのピボットテーブルを用いて、クロス集計表や基本統計量を示す表の作成方法を示します。
- Excelのピボットグラフを用いて、ピボットテーブルの情報をグラフで可視化する方法を紹介します。

【講座構成】

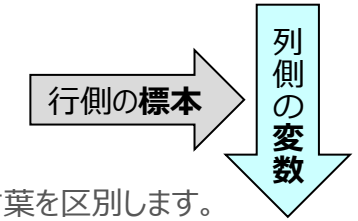
実習	[1] 変数の代表値としての基本統計量
	[2] 数値の尺度と代表値
	[3] 変数の代表値以外の基本統計量
	[4] ピボットテーブルによる表作成
	[5] ピボットグラフによる図作成

【学習のゴール】

- ✓ 標準的な基本統計量を理解し、Excel関数によって導出できる。
- ✓ Excelのピボットテーブルを用いて、クロス集計表、条件別の基本統計量の作成できる。
- ✓ Excelのピボットグラフを用いて、様々な観点からグラフによる可視化ができる。

1つの値で変数の特徴を示す「基本統計量」

◆基本統計量とは、変数の特徴をそれぞれ一つの値で要約する指標です。



- 一般に構造化データのデータセットは、個別の人・物・事象に対応する行側の**標本**と各標本に関する数量、属性を示す列側の**変数**で構成されます。
  - データという言葉は、「データセット全体」「標本」「変数」「ある標本のある変数」のいずれを指すかが不明瞭になりやすいため、言葉を区別します。
- 標本はサンプルとも呼ばれ、データセット内にある標本の数を**標本数**やサンプルサイズといいます。
  - 統計学において、「標本/サンプル」という言葉は「母集団」との対比においても使われ、本講座における「標本分散と標本標準偏差」でも言及します。
- 標本数が多くなると、各変数の標本を一つ一つ確認して、変数の特徴を把握することが困難になります。
- 平均値をはじめとして、変数の特徴を要約して、それぞれ一つの値で示す指標を**基本統計量**といいます。
  - 基本統計量は、記述統計量や要約統計量と呼ばれることもあります。
- 基本統計量には、最大値・最小値のように変数内の極端な特徴を示すものもあります。

データセット内の変数と標本の関係（［］内は例示）

	変数x ［身長(cm)］	**変数y** ［体重(kg)］	**変数z** ［性別：男性＝1、女性＝2］
標本1［Aさん］	150	40	2
標本2［Bさん］	160	50	1
標本3［Cさん］	170	70	1
標本4［Dさん］	180	90	1
標本5［Eさん］	170	60	2

変数x［身長(cm)］の基本統計量

- 平均値＝166
- 最大値＝180
- 最小値＝150

⋮

□ この講座の前半では、各基本統計量の意味やExcel関数を用いた基本統計量の導出を説明します。

実習用サンプルデータ（浜松市の気温と天気）

◆実習用サンプルデータとして2016年の静岡県浜松市の気温と天気のデータを利用します。

- 基本統計量やクロス集計表を例示する実習用サンプルデータとして、本講座では気象庁のウェブサイトからダウンロードした2016年1月1日～2016年12月31日の静岡県浜松市の気温・天気データを利用します。
  - 2016年（閏年により366日）における期間内の6時間毎の3時、9時、15時、21時の1日4時点の1464標本のデータセットを利用します。

気象庁のデータダウンロード用のウェブサイト

【出所】過去の気象データ・ダウンロード［気象庁］
http://www.data.jma.go.jp/gmd/risk/obsdl/index.php

ダウンロードデータを加工したExcelファイル

	A	B	C	D	E
1	年月日	月	時間	浜松の気温(℃)	浜松の天気
2	2016/1/1	01月	03時	5.6	晴れ
3	2016/1/1	01月	09時	7.7	晴れ
4	2016/1/1	01月	15時	11.5	晴れ
5	2016/1/1	01月	21時	7	晴れ
6	2016/1/2	01月	03時	4.4	晴れ
7	2016/1/2	01月	09時	8.3	晴れ
8	2016/1/2	01月	15時	13.2	晴れ
9	2016/1/2	01月	21時	9.1	晴れ
10	2016/1/3	01月	03時	8.5	晴れ
11	2016/1/3	01月	09時	9.6	晴れ
12	2016/1/3	01月	15時	14.6	曇
13	2016/1/3	01月	21時	11.4	晴れ
14	2016/1/4	01月	03時	8.5	晴れ
15	2016/1/4	01月	09時	11.7	晴れ
16	2016/1/4	01月	15時	16.1	晴れ
17	2016/1/4	01月	21時	13.5	晴れ

- 気温の変数はD列、天気の変数はE列
- 1日の中で3時、9時、15時、21時の4時点
- 2行目から1465行目が標本（データレコード）

■ この講座では「【実習用データ】ICT3-3_基本統計量・クロス集計表の作成.xlsx」を用いて実習を行います。
利用するExcelのシート番号は、各スライド右上の〔〕内に示します。http://www.soumu.go.jp/ict_skill/dc/ict_3_3data.zip

代表的な値を示す基本統計量（代表値）

◆平均値、中央値、最頻値は、それぞれ変数の代表的な値を表す基本統計量です。

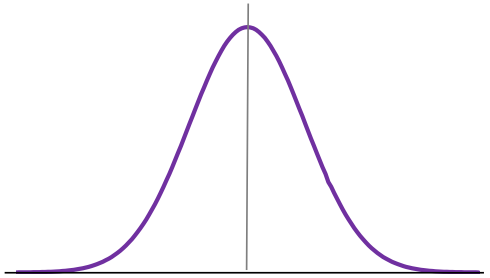
- **平均値（mean：ミーン）**は、標本の値の合計値を標本数で割った値に対応し、ExcelではAVERAGE関数を用いて『＝AVERGAGE(対象範囲)』で導出します。
  - 平均値は、日常生活でも利用頻度が多く、統計学に馴染みのない方も日常的に使う基本統計量です。
- **中央値（median：メディアン）**は、標本の値を大小関係で並べた際に中央の順位にある標本の値に対応し、ExcelではMEDIAN関数を用いて『＝MEDIAN(対象範囲)』で導出します。
- **最頻値（mode：モード）**は、その変数において最も同じ値が多く、頻度が高い値に対応し、ExcelではMODE関数を用いて『＝MODE(対象範囲)』で導出します。
- 平均値、中央値、最頻値は、総称として**代表値**とも呼ばれます。
- 統計学において有名な正規分布では「最頻値＝中央値＝平均値」で一致しますが、一般に平均値、中央値、最頻値の値はそれぞれ異なります。
  - 平均値は、極端に値が大きい/小さい標本の影響を受けやすい一方で、中央値と最頻値は極端な値の標本の影響を受けにくいという特徴があります。

実習用データ（浜松市の気温）に関する
平均値、中央値、最頻値の出力［シート〔2〕のI列］

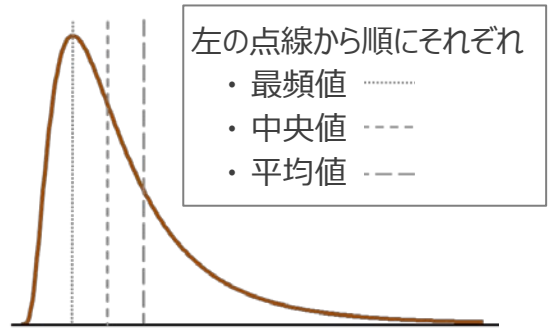
基本統計量	Excel関数の入力	出力
平均値	＝AVERAGE(D2:D1465)	17.7
中央値	＝MEDIAN(D2:D1465)	18.2
最頻値	＝MODE(D2:D1465)	24.7

平均値、中央値、最頻値のイメージ

正規分布の場合：
「最頻値＝中央値＝平均値」で一致



右側の裾が長い分布の場合：



代表値と数値の尺度

◆数値の尺度によって、代表値の各基本統計量が有用かどうかが異なります。

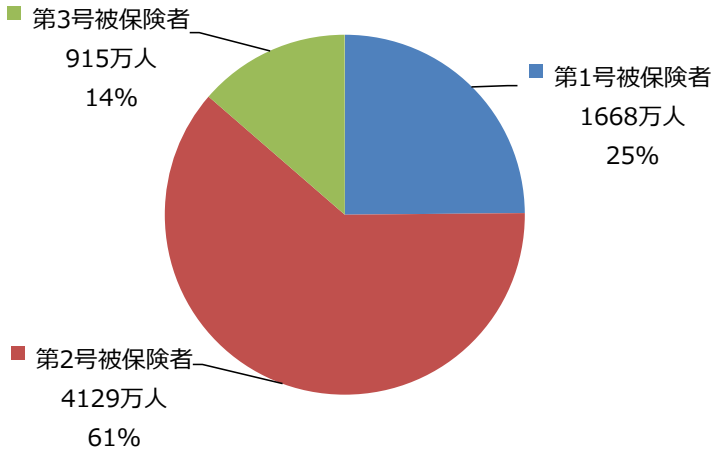
- 代表的な基本統計量である平均値においても、**数値の尺度**によって、無用で意味のない指標となることがあります。
- 日本における国民年金では、働き方等によって第1号、第2号、第3号と被保険者を区分しています。

国民年金の区分と被保険者数

区分	主な被保険者	人数（万人）
第1号被保険者	20歳以上60歳未満の自営業者・学生	1668
第2号被保険者	サラリーマン・公務員	4129
第3号被保険者	第2号被保険者の被扶養配偶者で20歳以上60歳未満の者	915

【出所】厚生年金保険・国民年金事業の概況（平成27年度）［厚生労働省］
http://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000106808_1.html

国民年金区分別の被保険者数の割合



- 例えば、国民年金における被保険者の区分の平均値を算出して、1.89号と導出しても意味がありません。

$$\text{被保険者の平均値} = \frac{1\text{号} \times 1668\text{万人} + 2\text{号} \times 4129\text{万人} + 3\text{号} \times 915\text{万人}}{6712\text{万人}} \fallingdotseq 1.89\text{号}$$

- **平均値には意味がない**一方で、『第2号被保険者の人数が最も多い』という**最頻値には意味があります**。

□ 続いて紹介する4種類の数値の尺度によって、それぞれの基本統計量が有用で意味があるかが異なってきます。

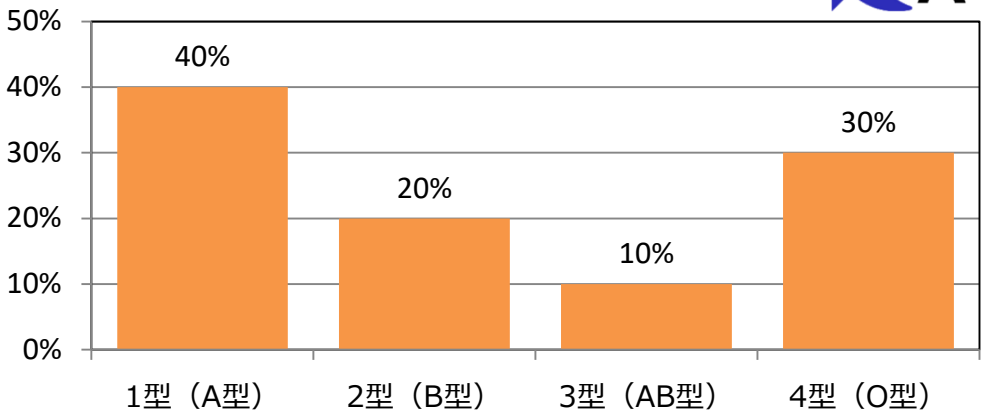
数値データの尺度（1）名義尺度 、順序尺度

◆一致か不一致のみに意味がある値を名義尺度、大小関係に意味があっても差や比率に意味がない値を順序尺度といいます。

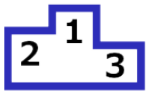
- 電話番号や郵便番号のように、区別や分類のみのために用いられる番号を**名義尺度**といいます。
  - 名義尺度では、一致する（等しい）かどうかのみに意味があり、大小関係に意味はありません。
- 郵便番号の数字をアルファベットに変更するように、名義尺度は、数値を記号に変換しても機能します。
  - 逆に記号から数値へ変更例として、血液型の「A型を1型、B型を2型、AB型を3型、O型を4型」と呼べば、名義尺度といえます。

- 地震の震度や5段階評価成績のように、数値に大小関係（順序）はあるものの数値の間隔に意味はないものを**順序尺度**といいます。
  - 5段階評価の5は、1～4よりも好成績とはいえますが、成績2と成績3を合わせた成績や、成績4の1.25倍の成績とはいえません。
  - 「震度3は震度2より揺れが強い」とはいえますが、「震度3は震度2と震度1が合わさった振動」や「震度3は震度2の1.5倍の揺れ」とはいえません。

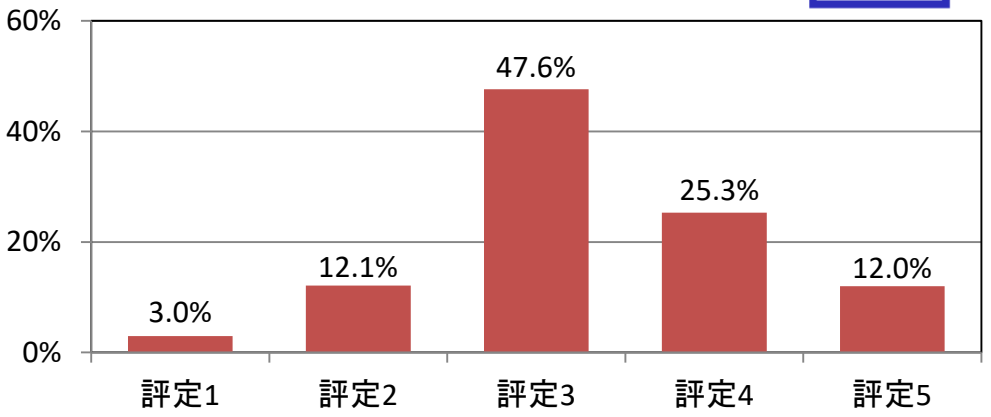
名義尺度の例（血液型）



【出所】日本人の血液型の発現率［日本赤十字社］に基づき作成
http://www.jrc.or.jp/donation/first/knowledge/



順序尺度の例（5段階の評点）

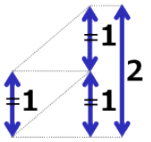


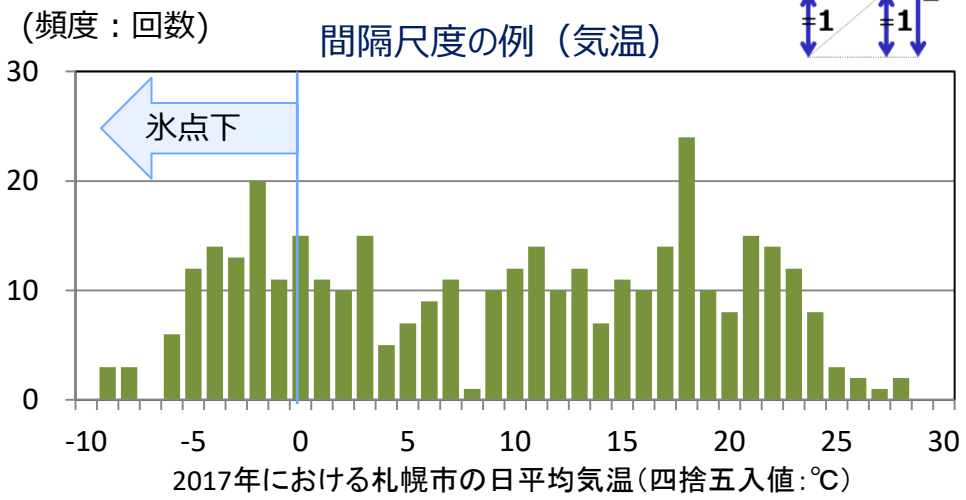
【出所】都内公立中学校第3学年（平成27年12月31日）の評定状況の調査結果［東京都］に基づき作成
http://www.metro.tokyo.jp/INET/CHOUSA/2016/03/60q3o200.htm

数値データの尺度（2）間隔尺度 、比率尺度

◆温度のように差分に意味はあっても比率に意味がない値を間隔尺度、身長のように差分と比率に意味がある数値を比率尺度といいます。

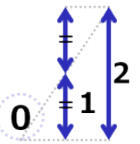
- 温度のように目盛が等間隔で差分に意味がある一方で、0や比率に意味がない数値を**間隔尺度**といいます。
  - 温度の「1℃→2℃」と「2℃から3℃」は「同じ1℃の上昇」とはいえますが、「3℃は1℃の3倍の温度」とはいえません。
  - 下図のように横軸に変数の範囲や項目を表し、縦軸に頻度（回数）を表すグラフを「ヒストグラム」といいます。

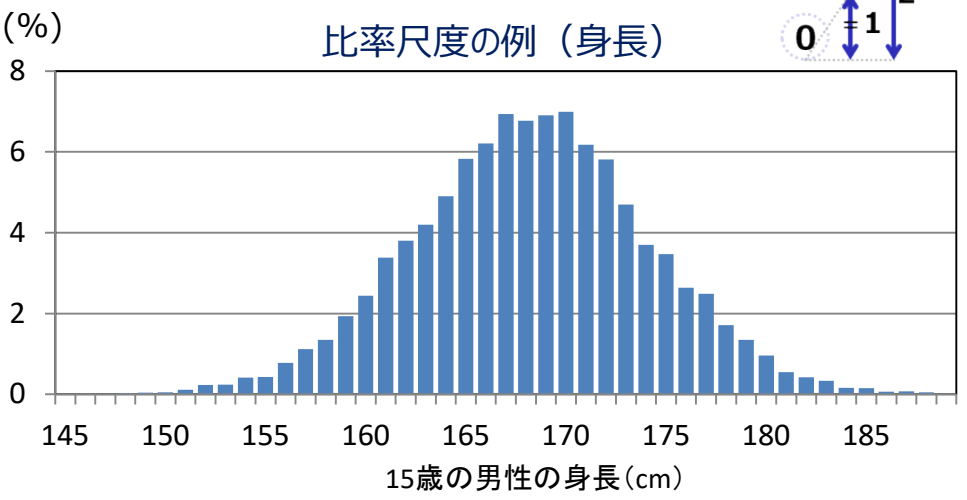




【出所】過去の気象データ・ダウンロード［気象庁］のデータに基づき作成
http://www.data.jma.go.jp/gmd/risk/obsdl/index.php

- 重量や長さのように、0に原点としての意味があり、間隔と比率の両方に意味がある尺度を**比率尺度**といいます。
  - 重量（g）や長さ（cm）では、「5gと10g」「1cmと2cm」の関係は、それぞれ2倍といえます。





【出所】平成28年度学校保健統計調査［文部科学省］に基づき作成
http://www.mext.go.jp/b_menu/toukei/chousa05/hoken/kekka/k_detail/1380547.htm

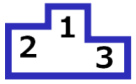
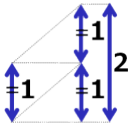
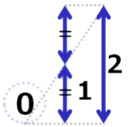
- 間隔尺度は0であっても「無や停止」に対応していません。一方、**比率尺度の0は「無や停止」に対応**します。
  - 0℃は水が凍る温度ではあっても、温度が存在しなくなるわけではありません。一方で重量0g、長さ0cmは、無に対応しています。

数値データの尺度と代表値のまとめ

◆数値データの尺度によって、有用な基本統計量が異なってきます。

- **数値データの尺度によって**、「大小比較」「差分計算」「比率計算」および**基本統計量の有用性が異なります**。

数値データの尺度と有用な演算・基本統計量の関係

尺度	事例	大小比較	差分計算	比率計算	代表的な値の表示に有用な基本統計量
名義尺度	郵便番号 部屋番号	×	×	×	• 最頻値
順序尺度 	震度 5段階評価	○	×	×	• 中央値 • 最頻値
間隔尺度 	温度（℃） 西暦	○	○	×	• 平均値 • 中央値 • 最頻値
比率尺度 	重さ 長さ	○	○	○	• 平均値 • 中央値 • 最頻値

- 郵便番号や部屋番号などの名義尺度は、最頻値のみが有用です。
- 5段階の評点などの順序尺度は、最頻値と中央値が有用です。
  - 満足度などの5段階評価の順序尺度においても、便宜的に平均値を指標として表すことがあります。しかし、同じ順序を保ったまま最上位の点数のラベルを5点から10点に変更すれば平均値は変化してしまうため、順序尺度に対して平均値は適切な指標とはいえません。
- 間隔尺度と比率尺度においては、平均値、中央値、最頻値の全てが有用です。

指定した順位の値を示す基本統計量

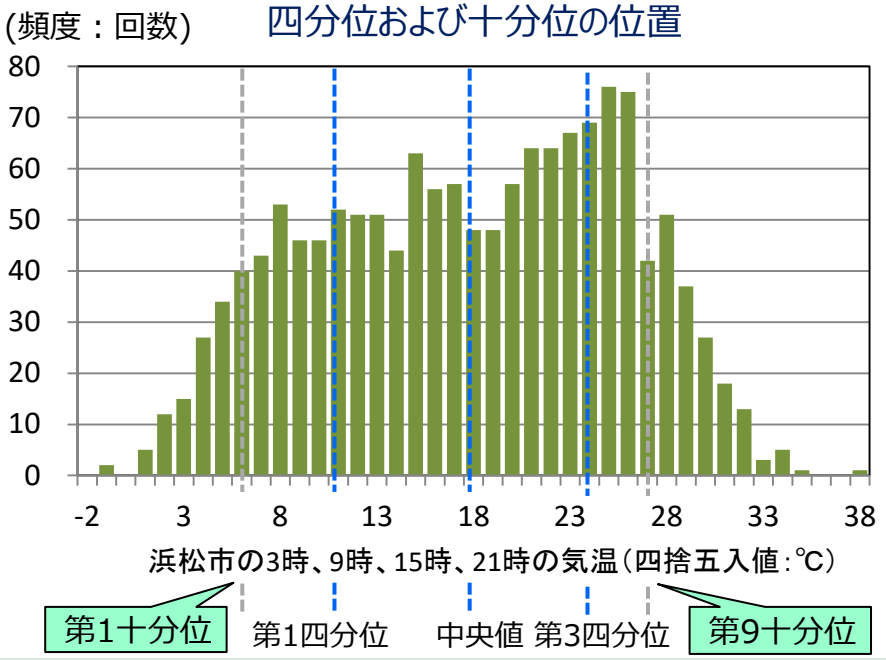
◆最大値、最小値、パーセンタイルは、変数内の特定の順位の値を示す基本統計量です

- **最大値**は変数の中で最も大きい値、**最小値**は変数の中で最も小さい値を示す基本統計量です。
- **パーセンタイル**は、変数を小さい方から数えて指定した割合における値を示す基本統計量です。
  - パーセンタイルは、小さい方から10%の値を示す第1十分位、25%の値を示す第1四分位、75%の値を示す第3四分位、90%の値を示す第9十分位で示すことが一般的です。
  - パーセンタイルの50%に該当する第2四分位、第5十分位は中央値（MEDIAN）でもあります。
- Excel関数を用いて、最大値は『=MAX(対象範囲)』、最小値は『=MIN(対象範囲)』、パーセンタイルは『=PERCENTILE(対象範囲,指定する小数値)』と、それぞれ記入することで出力が得られます。

実習用データ（浜松の気温）に関する
最大値、最小値、パーセンタイルの出力［シート〔2〕のM列］

基本統計量	Excel関数の入力	出力
最小値	=MIN(D2:D1465)	-3.0
第1十分位（＝小さい方から10%の値）	=PERCENTILE(D2:D1465,0.1)	6.8
第1四分位（＝小さい方から25%の値）	=PERCENTILE(D2:D1465,0.25)	11.3
第2四分位（＝中央値）	=PERCENTILE(D2:D1465,0.5)	18.2
第3四分位（＝小さい方から75%の値）	=PERCENTILE(D2:D1465,0.75)	24.3
第9十分位（＝小さい方から90%の値）	=PERCENTILE(D2:D1465,0.9)	27.6
最大値	=MAX(D2:D1465)	37.9

ヒストグラムにおける
四分位および十分位の位置



バラツキを表す基本統計量の導出

◆偏差および偏差平方和を経て、バラツキを表す基本統計量の分散と標準偏差を導出します。

- バラツキの指標を導出するために、まず各標本の標本平均からのズレとしての**偏差**（へんさ）を算出します。
  - 統計学における標準的な表記として、iは標本の順番を指し、x_iと下添え字で書くことによって変数xのi番目の標本を表します。
  - 標本数はnで表し、変数の合計値を標本数で割った値が標本平均です。変数xの標本平均は$\bar{x}$（エックスバー）と変数の上に横線を書いて表記します。

$$\text{変数}x\text{の}i\text{番目の標本} = x_i \qquad \text{変数}x\text{の標本平均} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad i\text{番目の標本の偏差} = x_i - \bar{x}$$

- 各標本の偏差を二乗することで負の偏差も全て正の値にしてから、全標本で総和をとった値が**偏差平方和**です。
  - 偏差には標本平均を中心に正と負の値が両方があり、二乗をせずに全標本で偏差の総和をとると、正と負が打ち消し合って0になってしまいます。
- 偏差平方和を標本数で割り、標本一つあたりのバラツキの大きさを表す基本統計量が（母集団としての）**分散**です。
  - 「母集団（ぼしゅうだん）」や「母集団としての」に関しては、次のスライドにて不偏分散（⇒標本分散）との対比で説明します。
- 分散は計算過程で尺度を二乗したバラツキの指標となっているため、分散の正の平方根をとることで尺度を元に戻したバラツキの大きさを表す基本統計量が（母集団としての）**標準偏差**です。
  - Excel関数として、偏差平方和は「DEVEQ」、（母集団としての）分散と標準偏差は、それぞれ「VAR.P」「STDEV.P」があります。
  - Excel関数において正の平方根を出力する「SQRT」もありますが、数値を0.5乗することによっても正の平方根を導出できます。

$$\text{偏差平方和} = \sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \text{分散} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \text{標準偏差} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

変数x［標本数(n)=3］に関するバラツキを示す指標の導出例［シート〔参〕の左側］

変数x		→	変数xの偏差	
標本1（x_1）	8	→	標本1（x_1）の偏差	-2
標本2（x_2）	10	→	標本2（x_2）の偏差	0
標本3（x_3）	12	→	標本3（x_3）の偏差	2

標本平均『=AVERAGE(対象範囲)』	10	=(8+10+12)/3
偏差平方和『=DEVSQ(対象範囲)』	8	=4+0+4
（母集団としての）分散『=VAR.P(対象範囲)』	2.667	=(4+0+4)/3
（母集団としての）標準偏差『=STDEV.P(対象範囲)』	1.633	=((4+0+4)/3)^0.5

標本分散と標本標準偏差

◆標本分散、標本標準偏差は、標本に基づく偏りのないバラツキの指標の推定値です。

- 分散と標準偏差には、標本に基づく母集団への偏りのない推定値として、**不偏分散（⇒標本分散）**や**不偏標準偏差（⇒標本標準偏差）**という基本統計量もあります。
  - 統計学において、母集団は利用できない標本を含めて考察の対象とする全体のデータ、標本は分析者が利用可能な一部のデータレコードに対応します。
- 不偏分散（⇒標本分散）および不偏標準偏差（⇒標本標準偏差）は、前スライドの分散・標準偏差の導出において、標本数(n)で割っていた部分をそれぞれ(n-1)で割ることによって導出します。
  - 1つしか標本がない場合は、標本にばらつける余地はなく、バラツキの指標は0になります。標本がばらつける余地は（n-1）に対応し、標本に基づいて偏りなく母集団のバラツキの指標を推定するためには、推定値が小さめに偏らないように（n-1）で割ります。
  - 標本分散と標本標準偏差はExcel関数においては、それぞれ「VAR.S」「STDEV.S」で導出できます。

$$\text{不偏分散（⇒標本分散）} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \qquad \text{不偏標準偏差（⇒標本標準偏差）} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

- 下記の表では、それぞれ標本数3、平均10の変数w、x、yで標本分散と標本標準偏差の関係を示しています。
  - 変数wは±1、変数xは±2、変数yは±3でばらついており、変数w、x、yの順でバラツキの指標である標本分散、標本標準偏差は大きくなります。
  - 変数w、x、yの標本標準偏差は、それぞれ1、2、3となっており、標本標準偏差は平均的なバラツキの大きさに対応しています。

3種の変数の標本分散、標本標準偏差の導出［シート〔参〕の右側］

	変数w	変数x	変数y
標本1	9	8	7
標本2	10	10	10
標本3	11	12	13
標本分散『=VAR.S(対象範囲)』	1	4	9
標本標準偏差『=STDEV.S(対象範囲)』	1	2	3

正の平方根

標本分散・標本標準偏差の算出と分布との対応

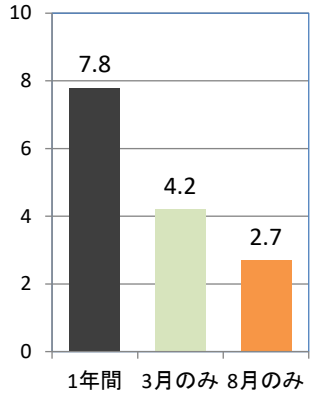
◆バラツキの大きさである分散・標準偏差の大きさは、分布の広がりに対応しています。

- ［浜松の気温(℃)］の標本分散と標本標準偏差を「1年間」「3月のみ」「8月のみ」で導出します。
  - 1年を通しての気温のバラツキ（分散・標準偏差）は、各1ヶ月の気温のバラツキよりも大きいことに加えて、季節の変わり目となる3月の方が、真夏の8月よりも気温のバラツキが大きいことが分かります。

3時、9時、15時、21時の気温データにおける分散と標準偏差のExcel出力［シート〔2〕のS列］

期間	標本数	基本統計量の名称	Excel関数の表記	出力
1年間	1464	（標本）分散	=VAR.S(D2:D1465z)	61.3
		（標本）標準偏差	=STDEV.S(D2:D1465)	7.8
3月のみ	123	（標本）分散	=VAR.S(D242:D365)	18.0
		（標本）標準偏差	=STDEV.S(D242:D365)	4.2
8月のみ	123	（標本）分散	=VAR.S(D854:D977)	7.2
		（標本）標準偏差	=STDEV.S(D854:D977)	2.7

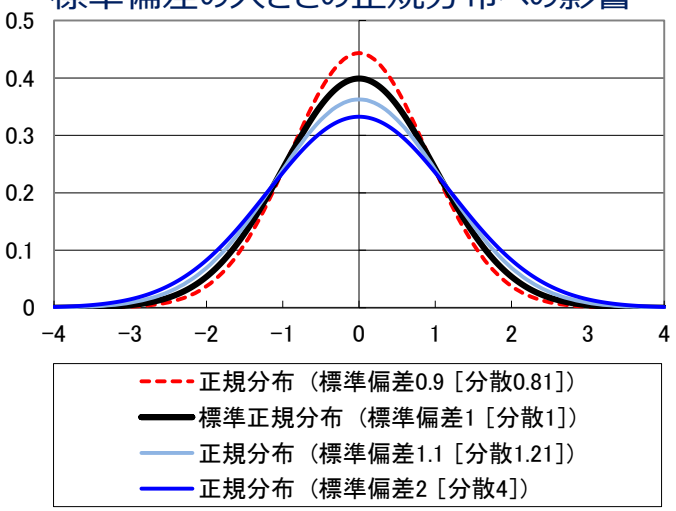
標準偏差の比較



- 同じ形状の分布において他の条件を固定して、分散・標準偏差を大きくすると、分布の頂点は低くなり、分布の裾は広がります。
- 統計学において有名な正規分布において、標準偏差と含まれる標本数の割合は下記のように対応しています。

平均値±1標準偏差	68.3%の標本が含まれる
平均値±2標準偏差	95.5%の標本が含まれる
平均値±3標準偏差	99.7%の標本が含まれる

標準偏差の大きさの正規分布への影響



ピボットテーブルによる集計

◆Excelのピボットテーブルを活用すると、簡単に集計表を作成することができます。

- 標本の件数を項目毎に集計した集計表には、**単純集計**と**クロス集計**があります。
  - オリジナルのデータセットから特定の2種類の変数（例：性別｜年齢層）で行と列を作り、項目（例：男性、女性|10歳代、20歳代、・・・）の交わる部分に該当する件数を求めることをクロス集計といい、クロス集計を表に表したものをクロス集計表といいます。

単純集計表の例（9時の天気）

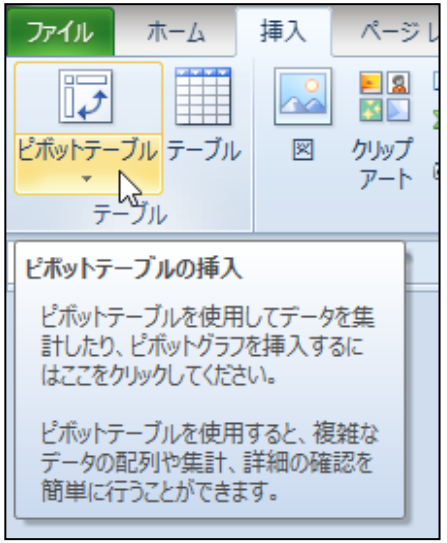
天気	頻度
晴れ	746
曇	543
雨	175
天気計	**1464**

クロス集計表の例（9時と21時の天気）

天気	9時	21時	時間数計
晴れ	746	780	**1526**
曇	543	564	**1107**
雨	175	120	**295**
天気計	**1464**	**1464**	**2928**

- Excelの「挿入」タブから選択できる**ピボットテーブル**を使うと、簡単に**単純集計表**や**クロス集計表**を作成できます。
  - ピボット（pivot）は、「中枢」や「旋回する（軸）」を表す英単語です。

ピボットテーブルのボタン



- ピボットテーブルでは、後述するように件数をカウントするクロス集計表のみならず、データセット内の項目ごとの合計値を示す表および平均値をはじめとする**基本統計量を示す表**の作成も可能です。
  - 講座3-2では、SUMIF関数による項目別合計値の導出、AVEREGEIF関数による項目別平均値の導出を示しましたが、ピボットテーブルにおいても同様に項目別の集計が可能です。

ピボットテーブルの作成

◆集計対象とするデータセットと配置先を指定して、ピボットテーブルを作成します。

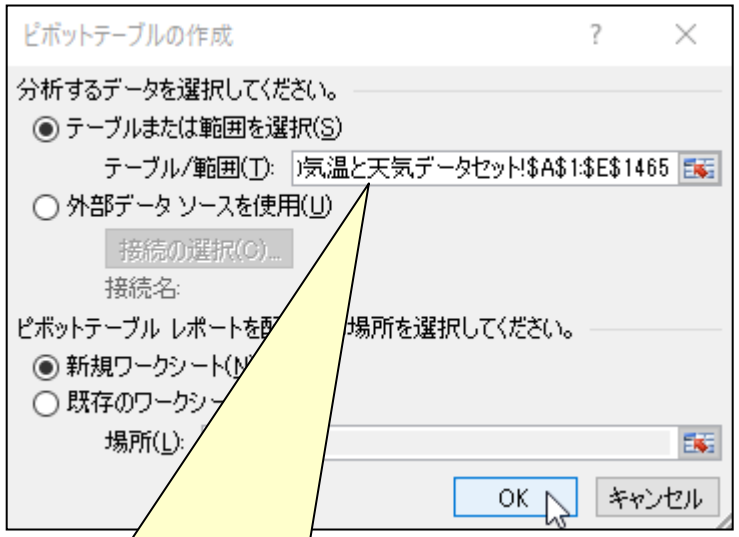
- 実習用データのデータセットを用いて、ピボットテーブルを作成します。

シート〔1〕のデータセットがある範囲を選択して、Excelの「挿入」タブにある「ピボットテーブルの作成」をクリックしてください。
データセットの「A1～A1465」の範囲指定を確認して、配置する場所は「新規ワークシート」として「OK」をクリックしてください。

- 実習用データのようにA列からデータセットが始まる場合は、対象とする範囲が初期設定として自動で入力されます。
- 配置先としてExcel内の新たなシートに集計表を作成する場合は「新規ワークシート」を選択する一方で、データセットと同じワークシート内に集計表を作成する場合は「既存のワークシート」を選択して、集計表の左上部分に当たるセルを指定します。

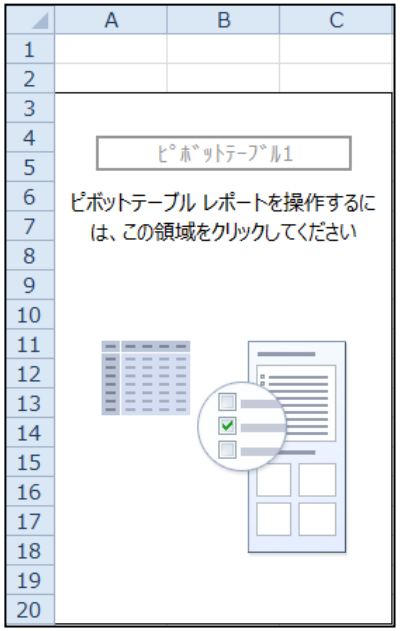
- 指定した配置先に「ピボットテーブルの枠」、Excelの右側に「ピボットテーブルのフィールドリスト」が表示されます。

「ピボッドテーブルの作成」のダイアログボックス

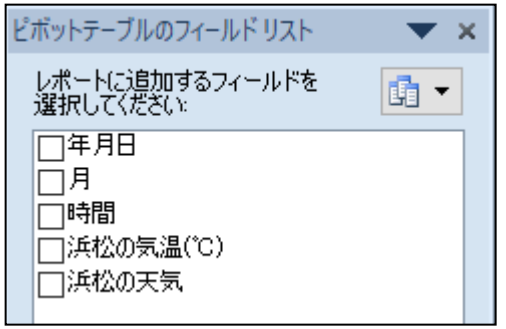


範囲が空白の場合や選択範囲が正しくない場合は、データセットの左上セルを選択してからキーボードの[Ctrl]と[Shift]を押しながら［→］、［↓］と押すとデータセット全体を選択できます。

ピボッドテーブルの枠



ピボッドテーブルのフィールドリスト



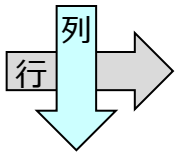
作成するクロス集計表の特定

◆ピボットテーブルを用いて作成するクロス集計表における行と列を決め、枠を作成します。

- ここでの例示ではクロス集計表における「横側の行」を［浜松の天気］とし、「縦側の列」を［月］として作成します。

作成する表形式：

	01月	02月	…	12月
晴れ	○回	○回	…	○回
曇	○回	○回	…	○回
雨	○回	○回	…	○回



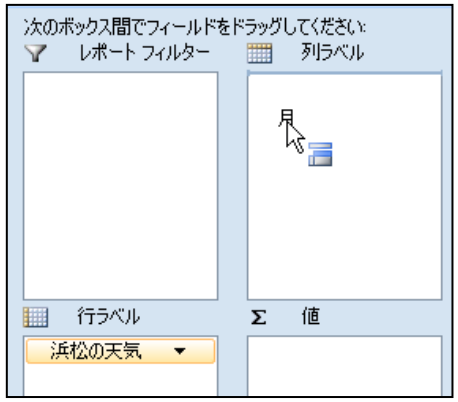
- 横側の行（ラベル）：浜松の天気
- 縦側の列（ラベル）：月

ピボットテーブルのフィールドリストから「横側の行」に入れる［浜松の天気］、「縦側の列」に入れる［月］をExcel画面の右上にあるフィールドリストから、右下側にある「行ラベル」と「列ラベル」ボックスへそれぞれドラッグ＆ドロップで移します。

- 二次元のクロス集計表を作りたい場合は、行ラベル、列ラベルの双方に区分となる変数を入力しますが、一次元の単純集計表を作る場合は、行ラベルのみに区分けをする変数を入力します。

- ピボットテーブルの**「行ラベル」**と**「列ラベル」**がついた枠がExcelのシート上に作成されます。
  - 「雨、晴れ、曇」の順となっているなど、作成したいラベルと順番が異なるケースもありますが、クロス集計表の作成後に順番を修正します。

フィールドリストから
各ラベルへドラッグ&ドロップ



表示されるピボットテーブルの行ラベルと列ラベル

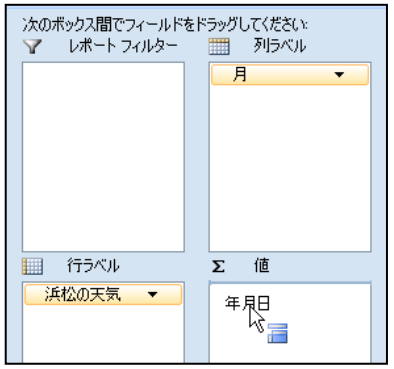
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2														
3		列ラベル												
4	行ラベル	10月	11月	12月	01月	02月	03月	04月	05月	06月	07月	08月	09月	総計
5	雨													
6	晴れ													
7	曇													
8	総計													

クロス集計表の作成と調整

◆ピボットテーブルを用いてクロス集計表を作成し、行・列のラベルの位置を整えます。

- Excelの画面右下の「値」の枠に［年月日］をドラッグ&ドロップして、クロス集計表を作成します。
  - 「値」に入力する項目は［年月日］でも［時間］でも構いませんが、連続的な数値データの入った［浜松の気温(℃)］を入力すると、初期設定における集計表の出力が「値の合計」になります。［年月日］［時間］のテキストデータであれば、初期設定で集計票の出力が「データの個数」となります。

「値」へドラッグ&ドロップ



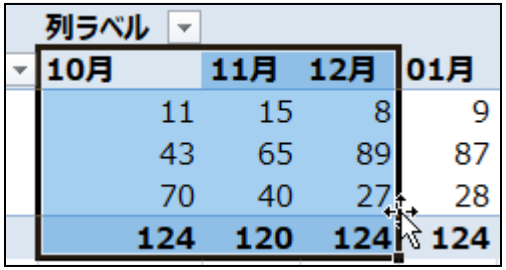
クロス集計表の初期表示

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2														
3	データの個数 / 年月日	列ラベル												
4	行ラベル	10月	11月	12月	01月	02月	03月	04月	05月	06月	07月	08月	09月	総計
5	雨	11	15	8	9	10	13	22	23	25	9	7	23	175
6	晴れ	43	65	89	87	81	76	46	56	36	53	75	39	746
7	曇	70	40	27	28	25	35	52	45	59	62	42	58	543
8	総計	124	120	124	124	116	124	120	124	120	124	124	120	1464

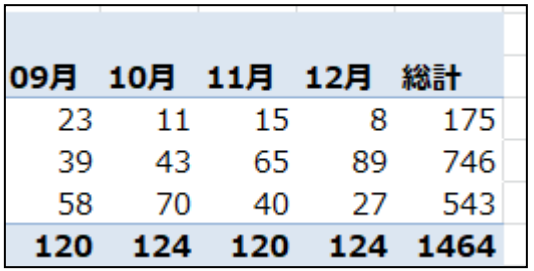
- ピボットテーブルの行・列の表示順を修正したい場合は、行・列を範囲で選択して移動させることができます。
  - Excelの機能でラベルを「昇順/降順」で並べ替えることができますが、独自の順序に並べ替えたい場合はドラッグ&ドロップが便利です。

［月］および［浜松の天気］に関して、移動させたい行・列を範囲を選択し、選択している枠の端をドラッグ&ドロップで移動させます。

［月］に関して移動させたい範囲の指定とドラッグ&ドロップ



［月］に関する列の位置移動



クロス集計表・ピボットテーブルの利用

◆ピボットテーブルのクロス集計表は頻度の把握に加え、一覧表示の基点にも利用できます。

- ピボットテーブルで作成した**クロス集計表**から、各変数の**項目の組み合わせの頻度**を把握できます。

行・列の位置を整理したクロス集計表

データの個数 / 年月日	列ラベル												
行ラベル	01月	02月	03月	04月	05月	06月	07月	08月	09月	10月	11月	12月	総計
晴れ	87	81	76	46	56	36	53	75	39	43	65	89	746
曇	28	25	35	52	45	59	62	42	58	70	40	27	543
雨	9	10	13	22	23	25	9	7	23	11	15	8	175
総計	**124**	**116**	**124**	**120**	**124**	**120**	**124**	**124**	**120**	**124**	**120**	**124**	**1464**

- ピボットテーブルにおいて、項目の組み合わせのセルをダブルクリックすると、組み合わせの条件に合致し、**集計対象となっている標本を新しいシートに一覧表示**します。
  - ピボットテーブルは、データ内の項目の組み合わせ別一覧を確認するための基点・インデックスとして利用することもできます。

標本が7件しかない「8月の雨」の標本を確認するために、ダブルクリックして該当する標本の一覧を表示します。

関心があるセルのダブルクリック

06月	07月	08月	09月
36	53	75	39
59	62	42	58
25	9	7	23
120	**124**	**124**	**120**

ダブルクリック

「8月の雨」の標本を抽出したリスト

	A	B	C	D	E
1	年月日	月	時間	浜松の気温(℃)	浜松の天気
2	2016/8/27	08月	21時	24.2	雨
3	2016/8/27	08月	15時	27.4	雨
4	2016/8/18	08月	09時	26.8	雨
5	2016/8/16	08月	03時	24.7	雨
6	2016/8/14	08月	03時	23.1	雨
7	2016/8/2	08月	21時	25.2	雨
8	2016/8/1	08月	15時	28.2	雨

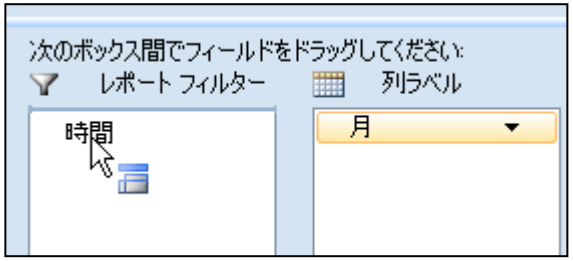
ピボットテーブルのフィルター（集計条件の設定）

◆ピボットテーブルでは、フィルターによって集計対象とする標本に条件をつけることができます。

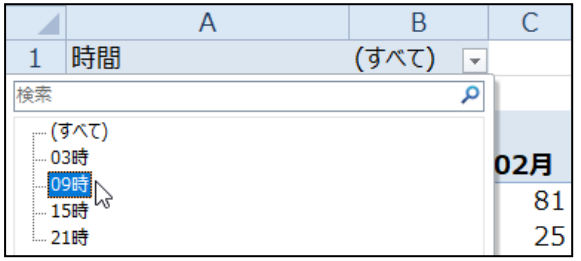
- Excelの右側の枠の**「レポートフィルター」**に変数を指定することで、集計対象とする標本に条件づけができます。

ピボットテーブルのフィールドリストから集計条件とする［時間］をレポートフィルターの枠へドラッグ＆ドロップで移します。
ExcelのA1・B1に表示されたフィルターから、絞り込みたい条件として「09時」を指定します。

レポートフィルターへのドラッグ＆ドロップ



フィルターに集計条件の［09時］の指定



- フィルターの「09時」の指定によって、午前9時のみを集計対象とした天気のクロス集計表が表示されます。
  - クロス集計表から、1月と12月は他の月に比べて「晴れの朝」が多く、『冬場の朝には晴れる傾向』があることが推察されます。

午前9時のみを集計対象としたクロス集計表

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	時間	09時												
2														
3	データの個数 / 年月日	列ラベル												
4	行ラベル	01月	02月	03月	04月	05月	06月	07月	08月	09月	10月	11月	12月	総計
5	晴れ	23	19	18	14	16	9	17	18	7	11	16	24	192
6	曇	5	8	10	11	11	15	13	12	17	17	12	5	136
7	雨	3	2	3	5	4	6	1	1	6	3	2	2	38
8	総計	31	29	31	30	31	30	31	31	30	31	30	31	366

1月や12月は、他の月に比べて「晴れの日」の頻度が高くなっています。

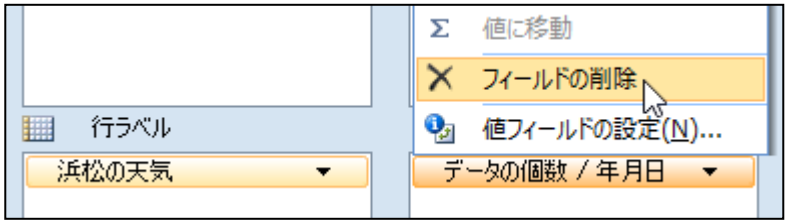
ピボットテーブルによる基本統計量の表作成

◆ピボットテーブルでは、変数の項目別に基本統計量を示す表を作成できます。

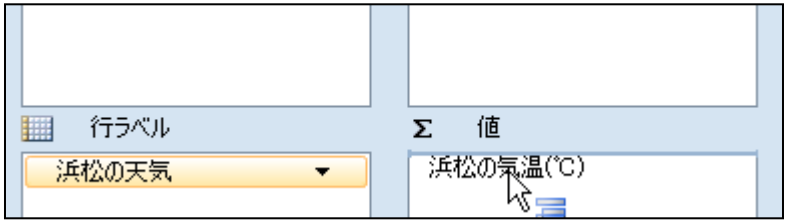
- **ピボットテーブル**は、項目別の頻度を表すクロス集計表以外にも、**基本統計量を示す表**を作成できます。

ピボットテーブルの「値」の欄にある［年月日］を左クリックして「フィールドの削除」をクリックします。
改めて、フィルードリストから、「値」の欄に［浜松の気温（℃）］をドラッグ＆ドロップします。

「値」に指定されていた［年月日］を削除



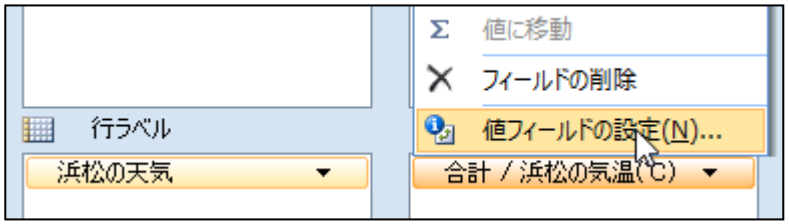
「値」に［浜松の気温（℃）］をドラッグ＆ドロップ



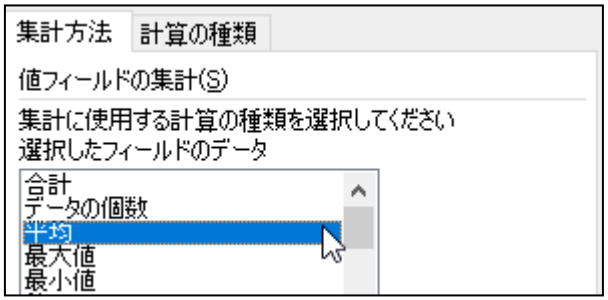
- **「値フィールドの設定」**を変更することで、項目別の平均値の表を作成します。
  - 気温のような連続的に変化する値をピボットテーブルの値として指定すると、初期設定では合計値を表示します。

「値」の欄にある「合計/浜松の気温（℃）」を左クリックして「値フィールドの設定」をクリックします。
表示されたダイアログボックスの「集計方法」のタブから「平均」を選択して「OK」をクリックします。

「値フィールドの設定」をクリック



「平均」を選択して「OK」をクリック



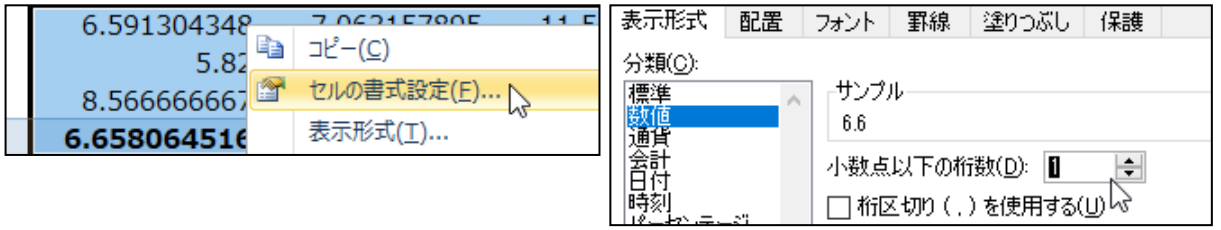
基本統計量の表（平均値）

◆ピボットテーブルによって項目別の平均値の表を見やすく表示し、出力を確認します。

- ピボットテーブルに表示された項目別平均値の小数表示や列幅を調整して、見やすい表に変更します。

対象範囲を指定し「セルの書式設定」から表示形式で「数値」を選択し、小数点以下の表示を調整して「OK」をクリックします。
対象列を指定して右クリックメニューの「列の幅」を選択して、列の幅を調整して「OK」をクリックします。

小数点以下の表示形式の調整



列幅の調整

- ピボットテーブルの表示を整え、午前9時における月別・天気別の平均気温を確認します。
- 1月、2月、12月においては「晴れの日」は「雨の日」よりも平均気温が低くなっており、『冬晴れの朝は寒くなる傾向』を把握できます。

午前9時における月別・天気別の平均気温

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	時間	09時												
2														
3	平均 / 浜松の気温(℃)	列ラベル												
4	行ラベル	01月	02月	03月	04月	05月	06月	07月	08月	09月	10月	11月	12月	総計
5	晴れ	6.6	7.1	11.6	18.5	22.3	24.6	28.8	29.8	28.7	22.8	15.1	9.5	17.1
6	曇	5.8	8.3	11.9	16.8	20.7	23.4	25.9	28.3	25.6	20.3	14.2	11.4	19.6
7	雨	8.6	12.6	10.9	15.0	19.3	21.7	20.0	26.8	24.5	20.6	10.3	15.8	17.7
8	総計	6.7	7.8	11.6	17.3	21.4	23.4	27.3	29.1	26.1	21.2	14.4	10.2	18.1

1月、2月、12月は、「晴れの日」は「雨の日」よりも平均気温が低いことが把握できます。

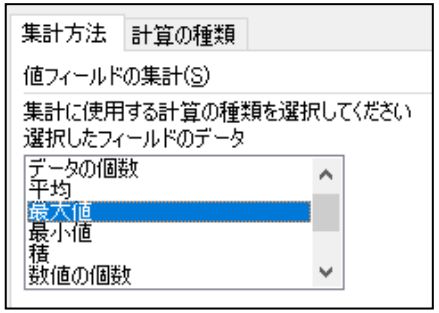
□ 後で紹介するピボットグラフでは、『冬晴れの朝は寒くなる傾向』をグラフによって可視化します。

基本統計量の表（最大値、標本標準偏差）

◆ピボットテーブルでは、最大値や標本標準偏差も項目別に表示できます。

- ピボットテーブルでは平均値以外にも、最大値や標本標準偏差といった基本統計量を表示することができます。
- ピボットテーブルから月別・天気別に午前9時の最高気温を確認することができます。

「値フィールドの設定」をクリックして、「集計方法」のタブから「最大値」を選択して「OK」をクリックします。

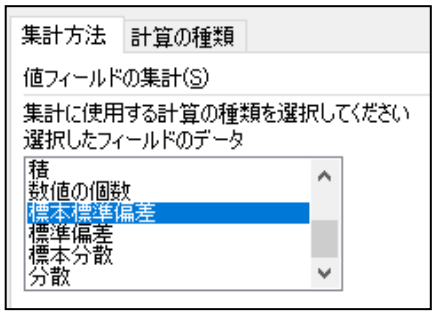


午前9時における月別・天気別の最高気温

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	時間	09時												
2														
3	最大値 / 浜松の気温(℃)	列ラベル												
4	行ラベル	01月	02月	03月	04月	05月	06月	07月	08月	09月	10月	11月	12月	総計
5	晴れ	11.9	10.6	16.8	21.2	25.1	27.9	32.3	32.7	30.4	27.0	19.4	14.0	32.7
6	曇	9.5	12.6	17.1	18.8	25.1	25.8	29.3	30.8	28.0	26.1	17.4	16.3	30.8
7	雨	9.7	16.7	14.9	17.1	20.8	24.5	20.0	26.8	26.7	22.0	12.7	17.3	26.8
8	総計	11.9	16.7	17.1	21.2	25.1	27.9	32.3	32.7	30.4	27.0	19.4	17.3	32.7

- ピボットテーブルから、月別・天気別にバラツキの指標である気温の標本標準偏差を確認することができます。
  - バラツキの指標は標本が1つしかないケースでは導出できず、計算過程で0で割ることになって導出不可を意味する「＃DIV/0!」が表示されます。
  - 集計対象を「全ての時間」とすると、Excel関数にて示した「1年間」「3月のみ」「8月のみ」の標本標準偏差と「総計」における合致を確認できます。

「値フィールドの設定」をクリックして、「集計方法」のタブから「標本標準偏差」を選択して「OK」をクリックします。



午前9時における月別・天気別の気温の標本標準偏差

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	時間	09時												
2														
3	標本標準偏差 / 浜松の気温(℃)	列ラベル												
4	行ラベル	01月	02月	03月	04月	05月	06月	07月	08月	09月	10月	11月	12月	総計
5	晴れ	3.0	2.2	3.9	1.8	1.5	2.1	1.6	1.1	1.1	2.2	3.1	2.3	8.9
6	曇	2.9	2.9	3.0	1.7	2.6	1.8	1.5	1.7	1.4	3.3	2.2	3.5	6.8
7	雨	1.3	5.8	3.7	2.7	1.2	2.1	#DIV/0!	#DIV/0!	1.5	2.3	3.5	2.2	5.8
8	総計	2.8	2.9	3.5	2.3	2.2	2.2	2.5	1.5	2.0	3.0	3.0	3.0	7.9

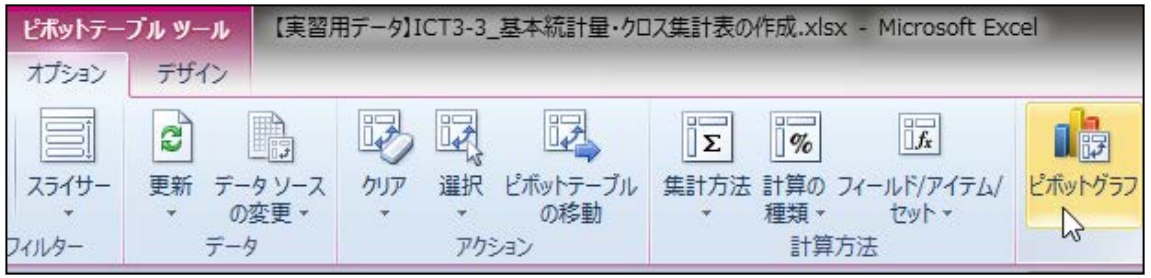
ピボットグラフの作成

◆ピボットグラフを利用すれば、ピボットテーブルの情報をグラフで可視化できます。

- **ピボットテーブル**に示された「午前9時における月別・天気別の平均気温」の情報を**グラフで可視化**します。

「値フィールドの設定」から「平均」に設定を戻し、表示されたピボットテーブル内のセルを選択します。
Excelの上部の「ピボットテーブルツール（オプション）」のタブに表示された「ピボットグラフ」をクリックします。
表示されたダイアログボックスの「グラフの挿入」から「集合縦棒」を選択して「OK」をクリックします。

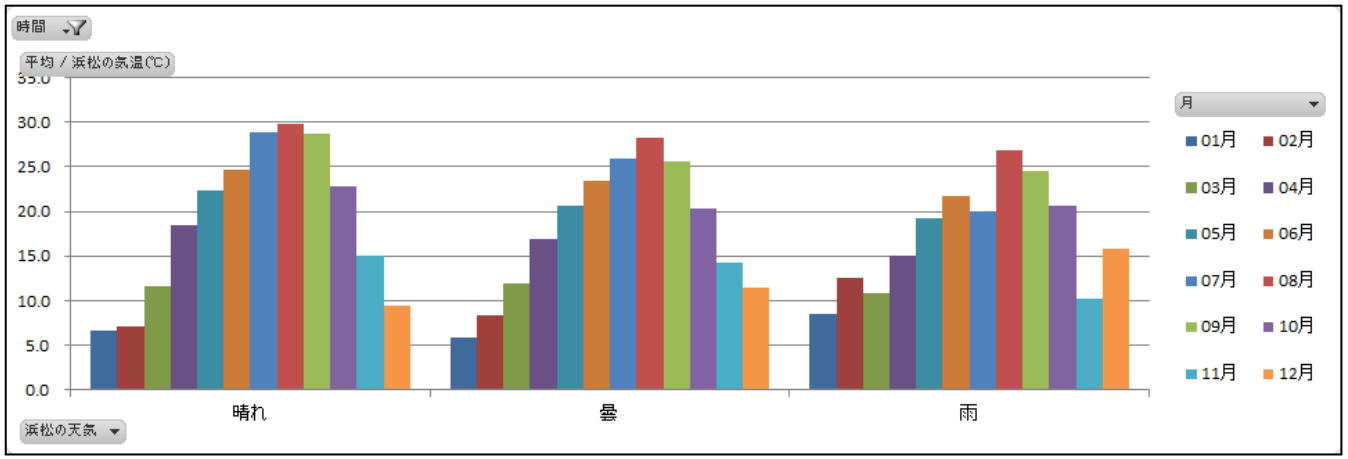
「ピボットテーブルツール（オプション）」のタブに表示された「ピボットグラフ」



グラフの指定

- **ピボットグラフ**によって「午前9時における月別・天気別の平均気温」をグラフで可視化することができます。

ピボットグラフによる「午前9時における月別・天気別の平均気温」のグラフ表示



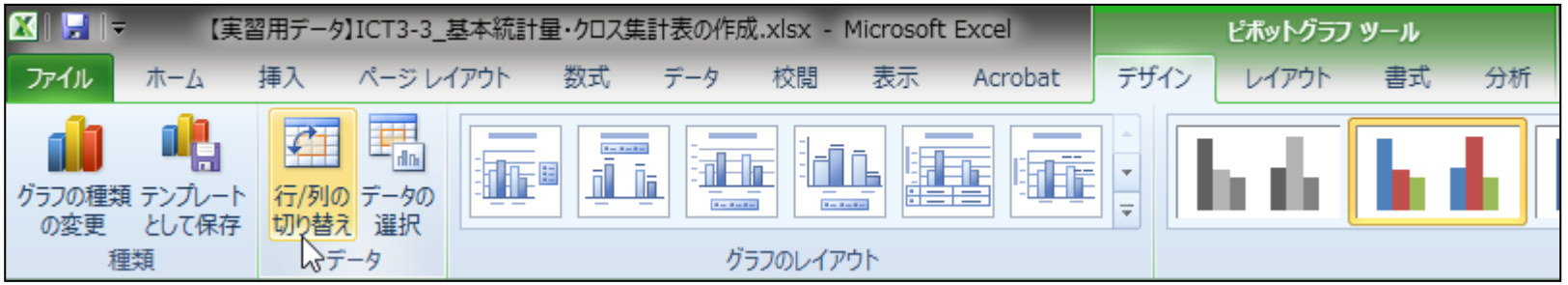
ピボットグラフの表示の切り替え

◆ピボットグラフの「行/列の切り替え」から、グラフの表示を切り替えることができます。

- ピボットテーブルで確認された『冬晴れの朝は寒くなる傾向』を明瞭に示すために、グラフの表示区分を切り替えます。

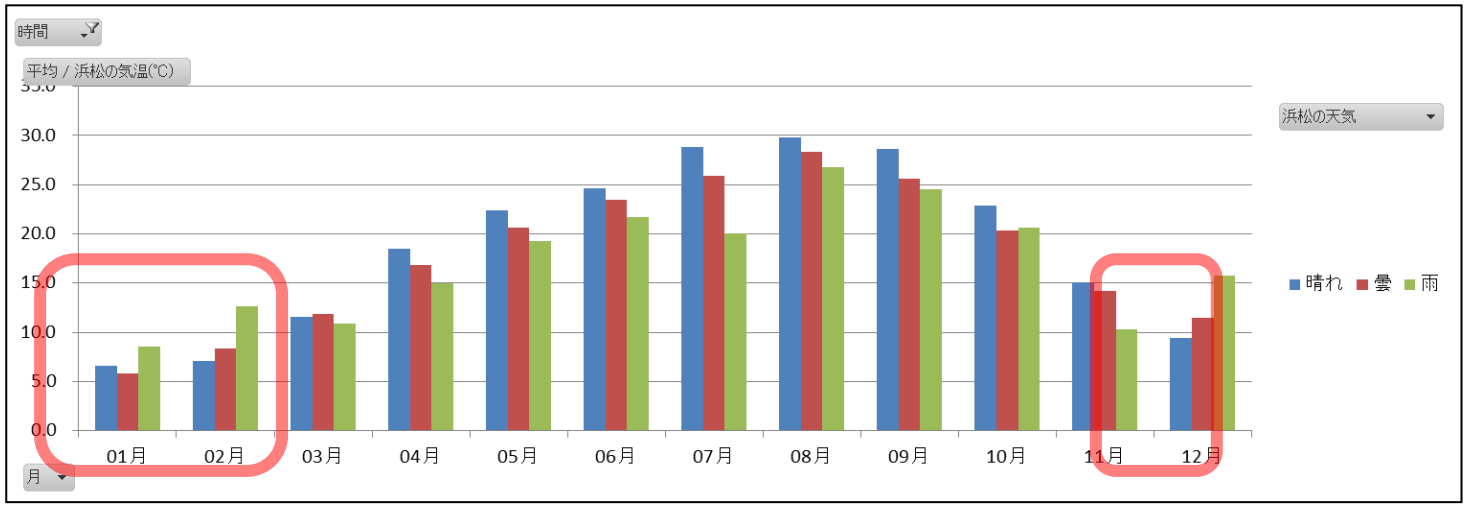
ピボットグラフを選択し、Excelの上部の「ピボットグラフツール（デザイン）」のタブに表示された「行/列の切り替え」をクリックします。

「ピボットグラフツール（デザイン）」のタブに表示された「行/列の切り替え」



- 表示を切り替えると、『冬晴れの朝は寒くなる傾向（雨の方が高気温）』を比較しやすいグラフで示すことができます。

表示を切り替えた「午前9時における月別・天気別の平均気温」のグラフ表示



スライサーの利用（表示対象の指定）

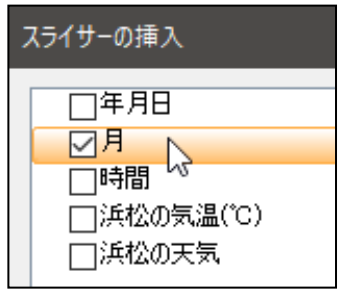
◆ピボットグラフのスライサーを利用すると、グラフの表示対象を簡単に指定することができます。

- 冬場に着目して『冬晴れの朝は寒くなる傾向』を示すために、ピボットグラフの機能の**スライサー**を利用します。

ピボットグラフを選択し、Excelの上部の「ピボットグラフツール（分析）」のタブに表示された「スライサー」を選択します。
「スライサーの挿入」のダイアログボックスから「月」にチェックを入れて「OK」をクリックします。

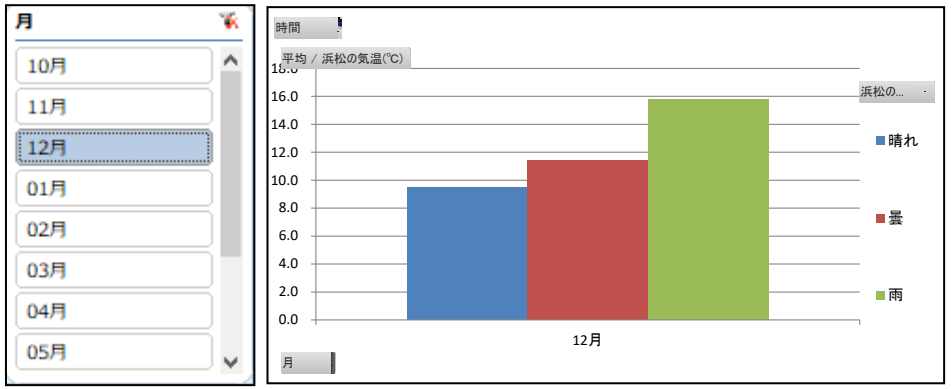
「ピボットグラフツール（分析）」のタブに表示された「スライサー」

「スライサー」の指定

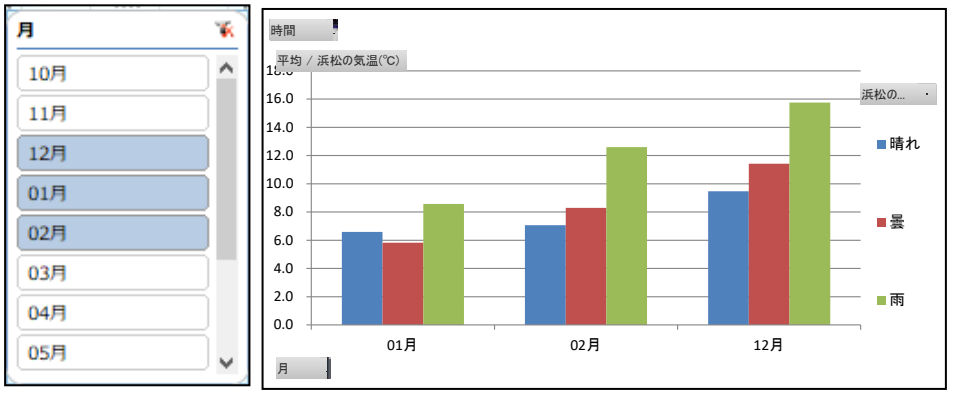


- 表示されたスライサーのボックスで表示したい［月］を指定することで、特定の月に**絞って表示**できます。
  - キーボードの「Ctrl」キーを押しながら、スライサーのボックス内のボタンをクリックすることで複数の月を選択して表示することができます。
  - ピボットグラフの左下に表示されている 月 ボタンからも同様の表示変更ができますが、スライサーの方がグラフへの反映が早く、比較が容易です。

スライサーによる12月に着目した表示



スライサーによる1月、2月、12月に着目した複数月の表示



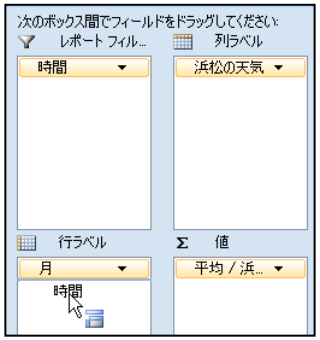
複数項目での行ラベルの指定

◆ピボットテーブル・ピボットグラフでは複数の項目を行ラベルに重ねて指定することができます。

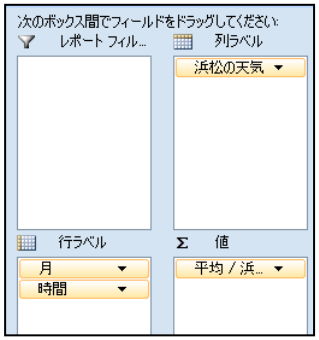
- 『冬晴れの朝は寒くなる傾向』が、昼や夜にも共通しているかを確認するために［時間］の観点を図表に加えます。
- ピボットテーブルを「午前9時」に限定するフィルターとして利用していた［時間］を**行ラベルの二段目**に移動させます。
  - ビボットグラフを選択すると、Excel右側の枠の表示はビボットテーブルにおける「行ラベル」が「軸フィールド」、「列ラベル」が「凡例フィールド」と変化します。

ピボットテーブルのセルを選択し「レポートフィルター」の枠に入っている「時間」を一つ下の枠の「行ラベル」にドラッグ&ドロップします。

行ラベルへの移動



2種の行ラベル

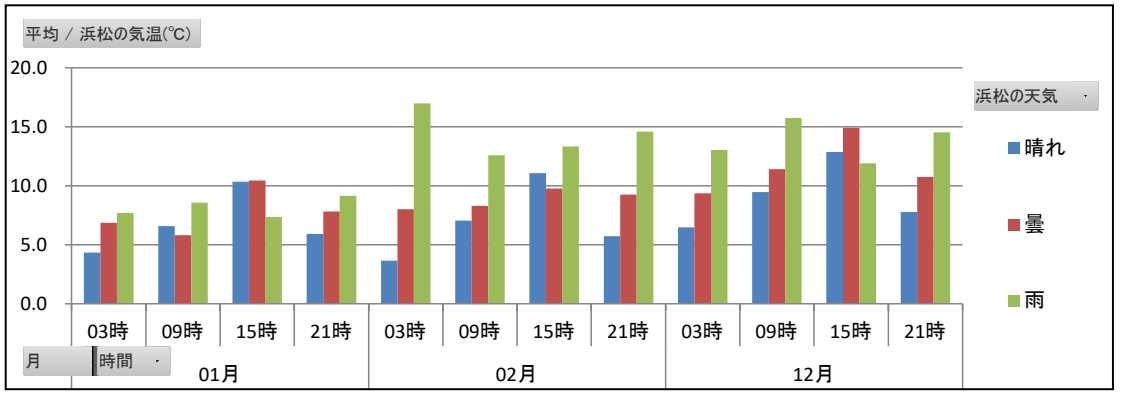


行ラベルに「月」「時間」を指定したビボットテーブル

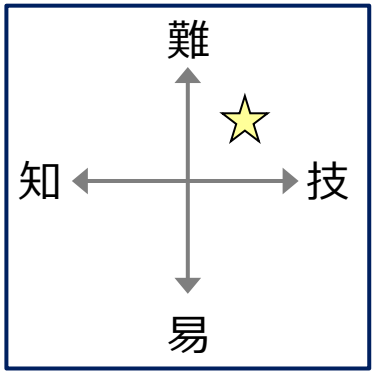
3	平均 / 浜松の気温(℃)	列ラベル			
4	行ラベル	晴れ	曇	雨	総計
5	**01月**	**6.7**	**8.0**	**8.2**	**7.1**
6	03時	4.3	6.9	7.7	5.1
7	09時	6.6	5.8	8.6	6.7
8	15時	10.3	10.5	7.4	10.2
9	21時	5.9	7.8	9.2	6.7
10	**02月**	**6.9**	**8.6**	**13.8**	**7.9**
11	03時	3.7	8.0	17.0	5.3

- 対応して表示が変化するビボットグラフの［月］のグラフの中に［時間］のグラフが現れ、**多角的な比較**ができます。
  - 15時に着目すると、1月と12月における晴れの平均気温は雨の平均気温を上回っており、朝と昼過ぎでは天気別の平均気温の傾向が異なっています。

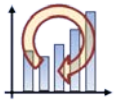
軸フィールドに「月」と「時間」を指定したグラフ



総務省　ICTスキル総合習得教材



［コース3］データ分析



3-4：相関と回帰分析（最小二乗法）

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_3_4.pdf

	1	2	3	4	5
［コース1］データ収集					
［コース2］データ蓄積					
［コース3］データ分析				●	
［コース4］データ利活用					

本講座の学習内容［3-4：相関と回帰分析（最小二乗法）］

【講座概要】

- Excelの散布図の作成方法、相関係数の導出方法、注意点を示します。
- 回帰分析（最小二乗法）の発想と用途を紹介します。
- Excelの分析ツールを用いた重回帰分析の実行方法を示します。
- Excelの分析ツールによる回帰分析の出力の直感的な意味を回帰分析全体と個別の説明変数に分けて説明します。

【講座構成】

実習

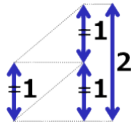
- [1] 散布図と相関係数
- [2] 回帰分析（最小二乗法）の発想
- [3] 分析ツールによる回帰分析
- [4] 回帰分析の全体に関する出力
- [5] 回帰分析の説明変数に関する出力

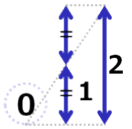
【学習のゴール】

- ✓ 散布図の表示、相関係数の導出方法、相関係数の利用上の注意点を把握する。
- ✓ 回帰分析（最小二乗法）の考え方を把握する。
- ✓ Excelの分析ツールを利用して、重回帰分析を行うことができる。
- ✓ Excelの回帰分析の出力の意味を理解する。

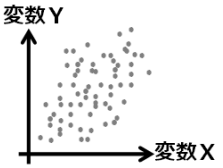
相関分析と回帰分析

◆この講座では、「相関分析」と「回帰分析」をExcelで実習します。

- この講座で説明する**「相関分析」**と**「回帰分析」**は、ともに連続的な数値をとる変数間の関係を分析する手法です。
- **連続的な数値**とは、講座3-3で示した**間隔尺度** 

または**比率尺度** 

に対応しています。
  - 名義尺度や順序尺度に相当する変数間の関係を分析する手法もありますが、標準的に利用される分析方法や手順の範疇外となります。
  - 来店客数などの人数は比率尺度ですが、「来店客が来るか？来ないか？」という来店客が0のケースがあり得るデータには、標準的な相関分析・回帰分析は適していません。一方で、来店客数の最小値が1人を上回り、来店客数の多寡を分析する場合には、相関分析・回帰分析が適しています。
- 「相関分析」と「回帰分析」は、ともに『**ある変数の大小が、他の変数の大小に結びついているか？』を分析する点では共通**していますが、**「分析者による因果関係の仮定」**や**「分析目的」に違い**があります。



相関分析 ⟷ **回帰分析**

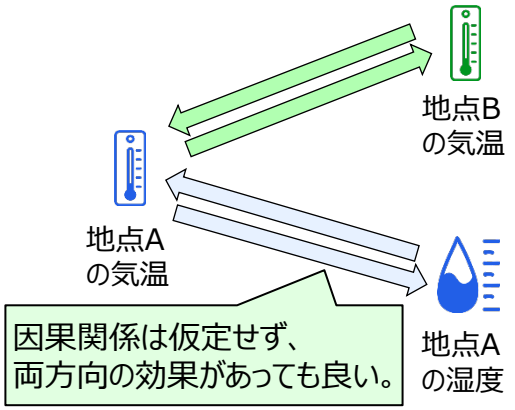
- 変数Xと変数Yの関係を分析する場合において、相関分析では因果関係を仮定しませんが、**回帰分析では「変数Xが変数Yに影響を与える」という「変数X→変数Y」の因果関係を分析者が仮定**した上で分析を行います。
- **相関分析では2種類の変数**を採りあげ、**変数間の（直線的な）関係の強さ**を測定します。
- 回帰分析の「変数X→変数Y」の因果関係において、**矢印の元になっている変数Xを説明変数**といい、**矢印の先になっている変数Yを被説明変数**といいます。
  - 回帰分析においては、説明変数Xには、XA、XB、XC…と複数の種類があっても同時に分析に入れることができます。
  - 回帰分析の説明変数は独立変数とも呼ばれ、被説明変数は従属変数や目的変数とも呼ばれます。
- 回帰分析では、**説明変数が被説明変数に与える効果の測定**を行ったり、**回帰分析の結果に基づいて設定した状況や将来に関する予測**を行ったりします。

相関分析と回帰分析が取り扱える変数の数

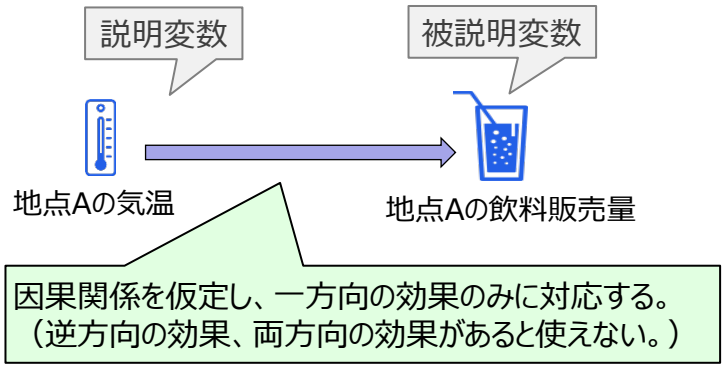
◆相関分析と回帰分析では、同時に取り扱うことができる変数の種類の数が異なります。

- **相関分析**は、**因果関係を仮定せず、常に一対一の組み合わせ**で変数同士の（直線的な）関係を分析します。
  - 相関行列として複数の相関係数をまとめて表示するケースもありますが、その要素となる各相関係数はそれぞれ一対一の変数の関係を示しています。
- **回帰分析**では、**分析者が説明変数（複数種も可）から被説明変数への因果関係を仮定**して分析を行います。
- **回帰分析**では**説明変数が1種の場合を単回帰分析**といい、**説明変数が2種以上の場合を重回帰分析**といいます。

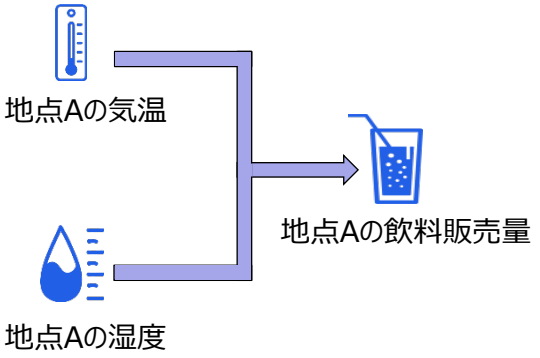
相関分析のイメージ



単回帰分析のイメージ



重回帰分析のイメージ



相関分析・回帰分析における仮定・分析目的・取り扱う変数の種類数の違い

分析の種類	分析者による因果関係の仮定	分析目的	取り扱う変数の種類数
相関分析	なし	• 変数間の（直線的な）関係の強さの測定	2種類（一対一）
回帰分析	あり	• 説明変数が被説明変数へ与える効果の測定 • 回帰分析の結果に基づく予測	2種類以上 （説明変数は複数種可）

■ この講座では「【実習用データ】ICT3-4_相関と回帰分析（最小二乗法）.xlsx」を用いて実習を行います。
利用するExcelのシート番号は、各スライド右上の〔〕内に示します。http://www.soumu.go.jp/ict_skill/dc/ict_3_4data.zip

Excelの「分析ツール」の導入

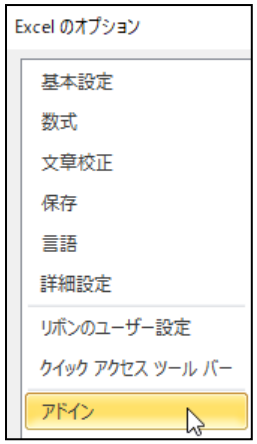
◆本講座ではExcelの分析ツールを利用するため、事前にアドインから導入しておきます。

- 相関分析・回帰分析の実習ではExcelの**分析ツール**を利用するため、未導入の場合は**アドインから導入**します。

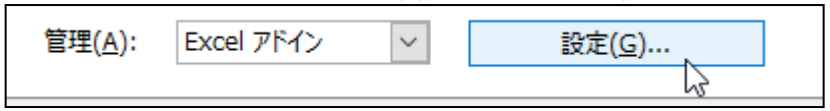
Excel画面の左上にある「ファイル」のボタンをクリックした後、「オプション」⇒「アドイン」をクリックして、「Excelアドイン」の設定ボタンをクリック後に「分析ツール」にチェックを入れて「OK」をクリックします。

- Excelのバージョンによって、分析ツールのアドインを呼び出す項目が異なります。ここではWindows用のExcel 2010で例示をしています。
- Macintosh用の「Excel for Mac 2011」以前では分析ツールのアドインはありません。「Excel for Mac 2016」から、Macintosh用のExcelにおいても分析ツールのアドインが利用できるようになりました。

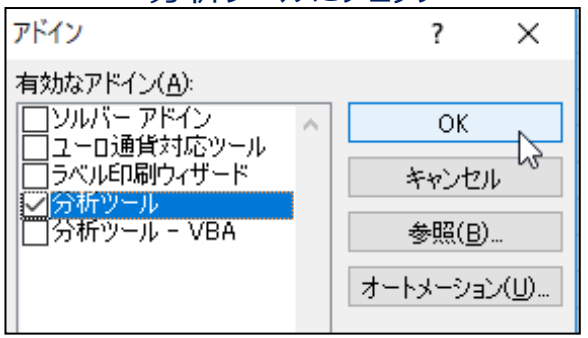
アドインのクリック



Excelアドインの設定ボタンをクリック

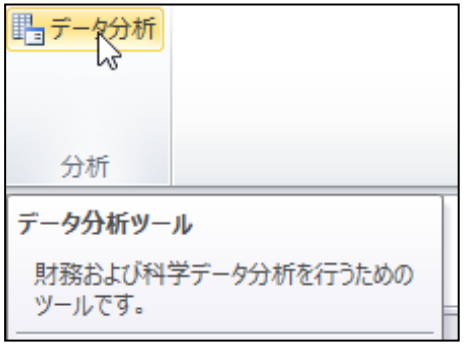


分析ツールにチェック

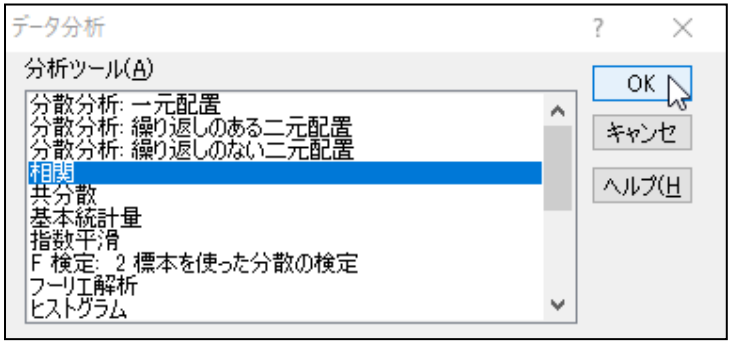


「分析ツール」の導入後、Excel画面の上側の「データ」タブに現れた「データ分析」をクリックすることで「データ分析」のダイアログボックスを呼び出すことができます。

データ分析をクリック



データ分析のダイアログボックス



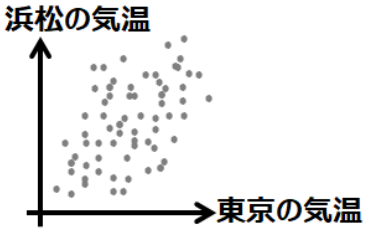
講座内の2箇所でExcelの「分析ツール」を利用して分析を行います。

散布図の作成

◆実習用データから2種類の変数を指定して散布図を描きます。

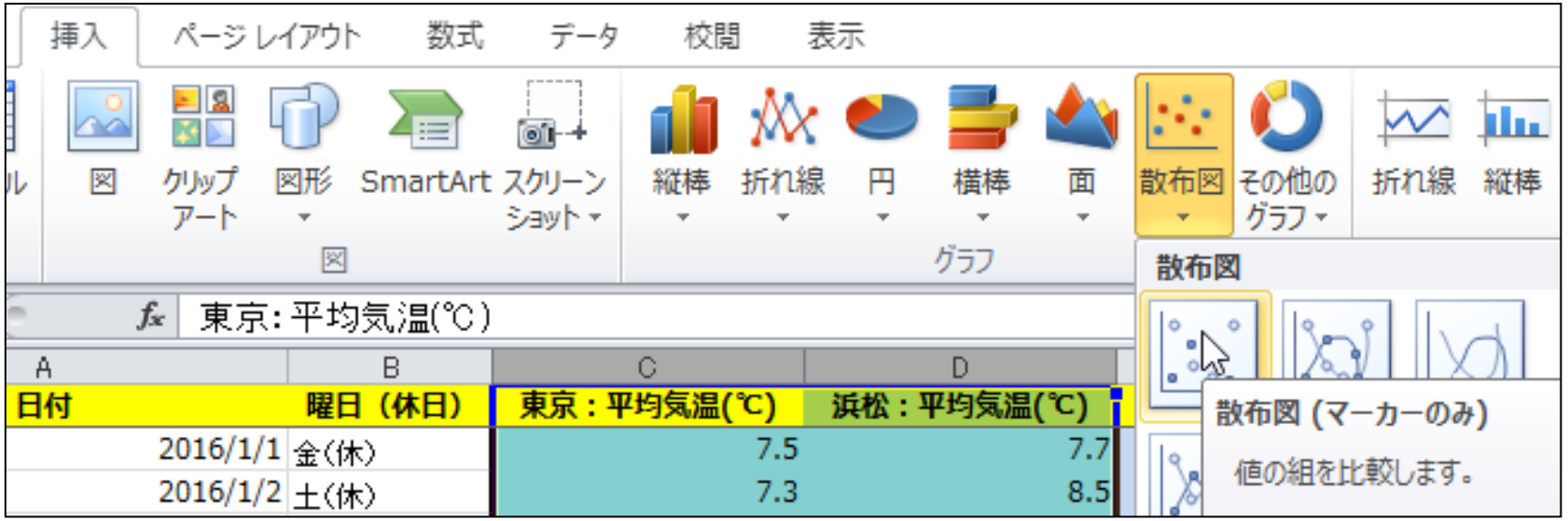
- 実習用データのシート〔1〕には、東京と浜松における日平均気温・湿度および飲料販売量が記入されています。
  - 気温、湿度のデータは気象庁のウェブサイトからダウンロードした2016年に関する現実のデータ、飲料販売量は架空のデータです。
    【出所】過去の気象データ・ダウンロード［気象庁］ http://www.data.jma.go.jp/gmd/risk/obsdl/index.php

- ここでは散布図作成の例示するために［東京：平均気温(℃)］と［浜松：平均気温(℃) ］の2変数の関係を描くことを考えます。



散布図の対象データとする2列をまとめて選択した状態で、
Excelの上部の「挿入」タブにある「散布図」ボタンをクリックし、その中にある「マーカーのみ」のボタンをクリックします。

「挿入」タブにある「散布図」から「散布図（マーカーのみ） のボタン」をクリック



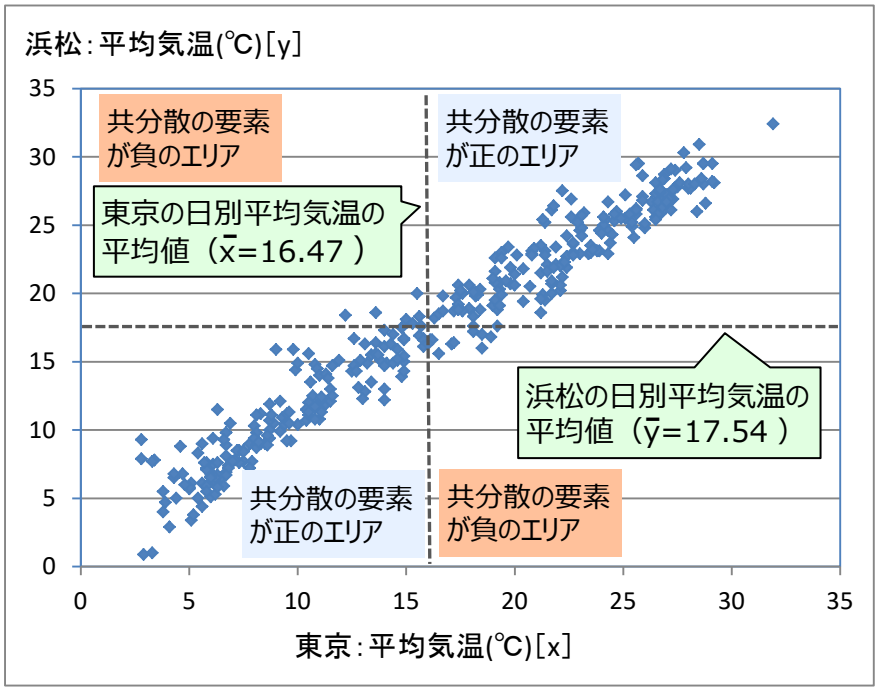
散布図と相関係数

◆相関係数は、2種類の連続的な変数の関係を一つの値で要約します。

- 作成した散布図は右図のように、東京の気温が高い場合に浜松の気温も高い傾向があり、直線的な関係が確認できます。
- 一方の変数が大きいと、もう一方の変数が大きい傾向があり、右肩上がりの直線的な関係が確認できる場合は**正の相関関係**にあるといいます。
  - 直線的な関係の傾きが右肩下がりの場合は負の相関関係にあるといい、直線的な関係が真横であったり、直線的な関係が確認しにくい場合は無相関といいます。
- 相関関係を「-1～1」の範囲で一つの値で要約する指標が、**相関係数**であり、下記の式で導出できます。
  - 右図では東京の気温をx、浜松を気温をyと取り扱っています。

$$x\text{と}y\text{の相関係数} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

東京と浜松の気温の散布図（2016年：日別平均）



- 相関係数の分子は、変数xと変数yの**共分散**（きょうぶんさん）と呼ばれます。共分散は各標本の「『変数xの偏差』と『変数yの偏差』の積」を要素として、標本での総和をとってから標本数で割ること（⇒平均値）で算出できます。
  - 右上のグラフにおいて、それぞれの平均値を軸として、右上と左下のエリアの要素は正の値となり、左上、右下のエリアの要素は負の値となります。
- 相関係数の分母は変数x、変数yの各標準偏差となっており、相関係数を「-1～1」の間に調整する役割を持ちます。
  - 標準偏差は、講座3-3で示したように変数のバラツキの指標として、二乗した値の総和によって導出しているので、必ず正の値になります。
- 相関係数は、**正と負の値をとりうる共分散が分子、正の値のみをとる「-1～1」への調整項が分母**になっています。

Excel関数CORRELによる相関係数の導出

◆Excel関数CORRELで2種類の変数を指定して相関係数を導出することができます。

- 相関係数の分子の**共分散は、各変数の偏差の積を要素とする平均値**となっていることから、散布図における各変数の平均値の軸から「右上」と「左下」が多ければ正の値、「右下」と「左上」が多ければ負の値になります。
- Excel関数「COVARIANCE.P」を利用すると、（母集団としての）共分散が導出できます。
  - 講座3-3において分散や標準偏差で示したように、標本数(n)ではなく(n-1)で割ることで標本共分散を導出するExcel関数「COVARIANCE.S」もあります。なお、分子を標本共分散とする場合は、対応する分母も標本標準偏差となり、分子分母をともに(n-1)で割るため、相関係数は一致します。
- 変数x、変数yのそれぞれの（母集団としての）標準偏差は、講座3-3で示したように「STDEV.P」で導出できます。
- 変数xと変数yの共分散をそれぞれの変数の標準偏差で割ることによっても、Excel関数「CORREL」を利用することによっても、変数xと変数yの相関係数を導出することができます。
- ［東京：平均気温(℃)］と［浜松：平均気温(℃) ］の相関係数として0.977が導出できます。

実習用データ（浜松の気温・東京の気温）に関する
共分散、相関係数の出力［シート〔2〕のL列］

指標	Excel関数の入力	出力
xの平均値	=AVERAGE(C2:C367)	16.47
yの平均値	=AVERAGE(D2:D367)	17.54
xとyの共分散	=COVARIANCE.P(C2:C367,D2:D367)	55.997
xの標準偏差	=STDEV.P(C2:C367)	7.668
yの標準偏差	=STDEV.P(D2:D367)	7.474
xとyの相関係数	=CORREL(C2:C367,D2:D367)	0.977

共分散を分子、
各標準偏差の積を分母
とする割り算

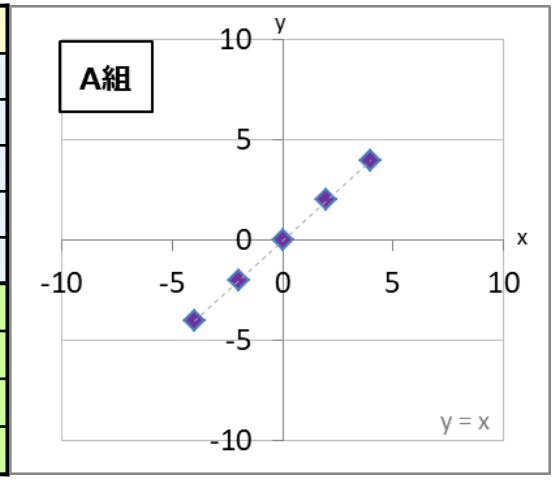
□ 続く2枚のスライドでは、5つの標本での簡潔な数値例で、相関係数の特徴を紹介します。

【参考】相関係数と散布図の例示（1）

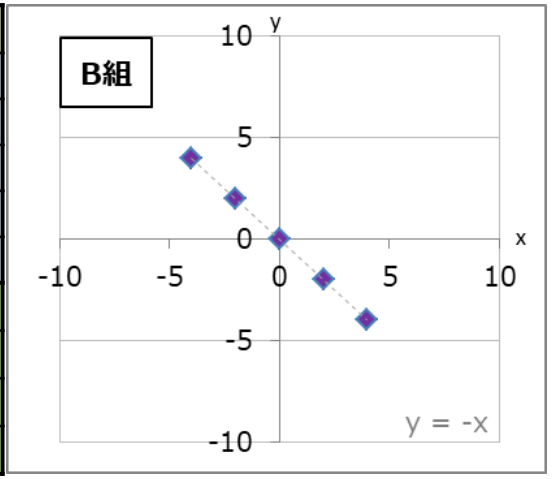
◆変数間の関係を示す直線上に全標本が乗っている場合、相関係数は絶対値で1になります。

- 相関係数は、標本全てが右肩上がりの直線上にあれば+1、標本全てが右肩下がりの直線上にあれば-1となります。
  - 散布図の右下には、後述する「散布図における単回帰分析」を行った場合に表示される回帰線の数式を示しています。

A組	変数x	変数y
標本1	-4	-4
標本2	-2	-2
標本3	0	0
標本4	2	2
標本5	4	4
xとyの共分散	8.000	
xの標準偏差	2.828	
yの標準偏差	2.828	
xとyの相関係数	**1.000**	

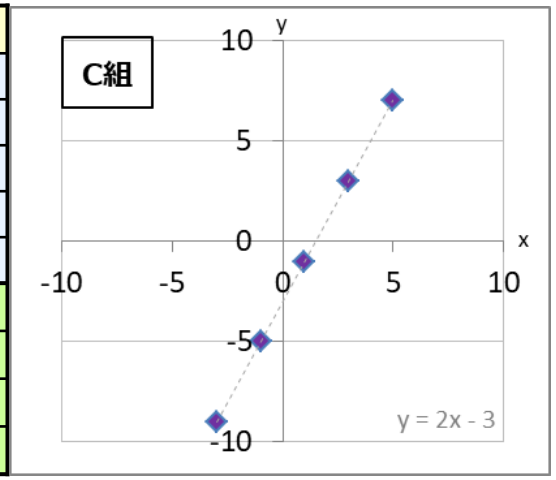


B組	変数x	変数y
標本1	-4	4
標本2	-2	2
標本3	0	0
標本4	2	-2
標本5	4	-4
xとyの共分散	-8.000	
xの標準偏差	2.828	
yの標準偏差	2.828	
xとyの相関係数	**-1.000**	

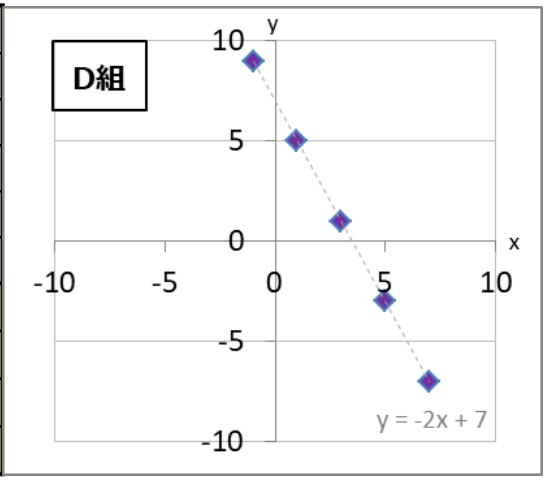


- 標本を平行移動させても、（全標本が直線上にある場合は）直線の傾きが変わっても、相関係数は変わりません。

C組	変数x	変数y
標本1	-3	-9
標本2	-1	-5
標本3	1	-1
標本4	3	3
標本5	5	7
xとyの共分散	16.000	
xの標準偏差	2.828	
yの標準偏差	5.657	
xとyの相関係数	**1.000**	



D組	変数x	変数y
標本1	-1	9
標本2	1	5
標本3	3	1
標本4	5	-3
標本5	7	-7
xとyの共分散	-16.000	
xの標準偏差	2.828	
yの標準偏差	5.657	
xとyの相関係数	**-1.000**	

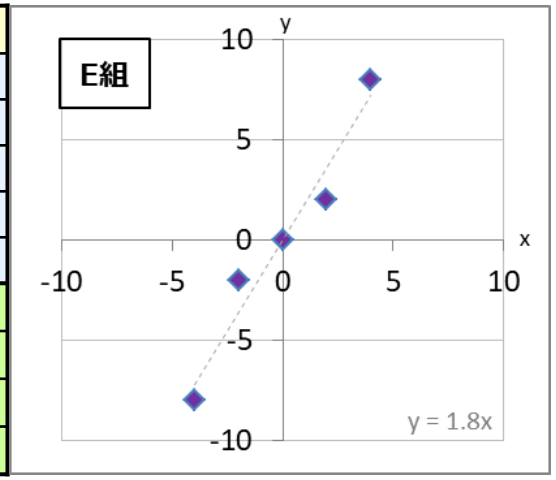


【参考】相関係数と散布図の例示（2）

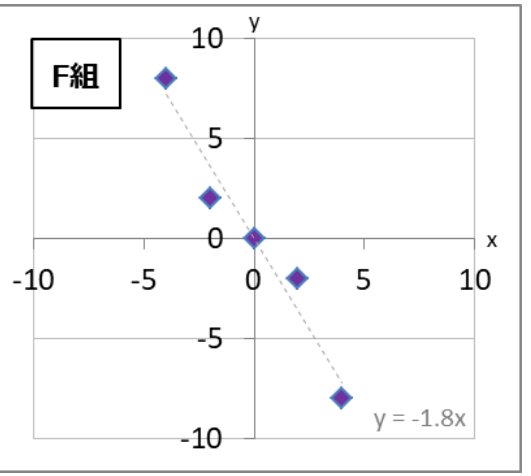
◆変数間の関係を示す直線から外れる標本があれば、相関係数は絶対値で1を下回ります。

- 変数間の変化が一方向でも、直線的な関係から外れる標本があれば、相関係数は絶対値で1を下回ります。

E組	変数x	変数y
標本1	-4	-8
標本2	-2	-2
標本3	0	0
標本4	2	2
標本5	4	8
xとyの共分散	14.400	
xの標準偏差	2.828	
yの標準偏差	5.215	
xとyの相関係数	**0.976**	

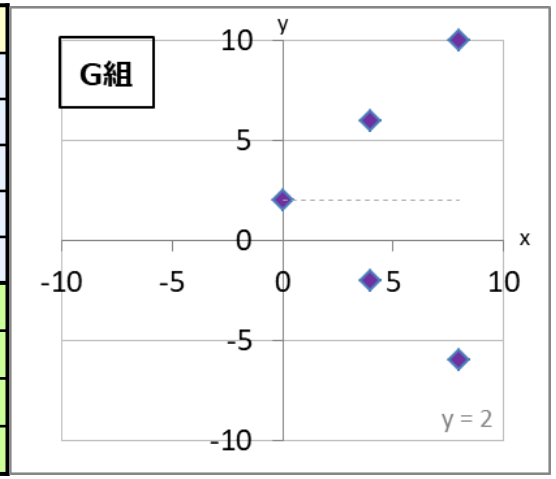


F組	変数x	変数y
標本1	-4	8
標本2	-2	2
標本3	0	0
標本4	2	-2
標本5	4	-8
xとyの共分散	-14.400	
xの標準偏差	2.828	
yの標準偏差	5.215	
xとyの相関係数	**-0.976**	

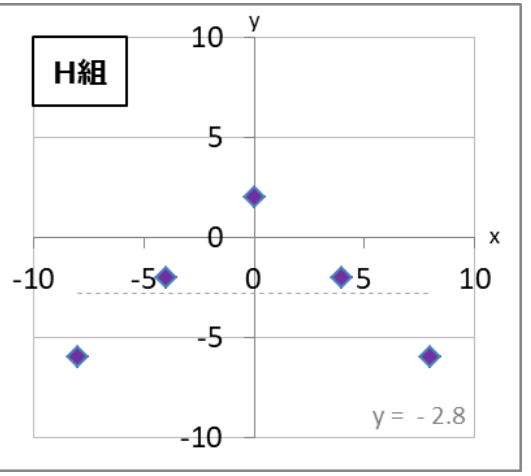


- 相関係数は屈曲した関係を反映せず、2種類の変数間に関係があっても、相関係数が0になるケースもあります。

G組	変数x	変数y
標本1	8	10
標本2	4	6
標本3	0	2
標本4	4	-2
標本5	8	-6
xとyの共分散	0.000	
xの標準偏差	2.993	
yの標準偏差	5.657	
xとyの相関係数	**0.000**	



H組	変数x	変数y
標本1	-8	-6
標本2	-4	-2
標本3	0	2
標本4	4	-2
標本5	8	-6
xとyの共分散	0.000	
xの標準偏差	5.657	
yの標準偏差	2.993	
xとyの相関係数	**0.000**	



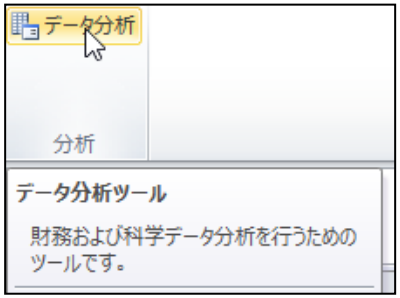
- シート〔参1〕では、散布図における各標本の位置を変更して、相関係数がどのように変化するかを確認できます。

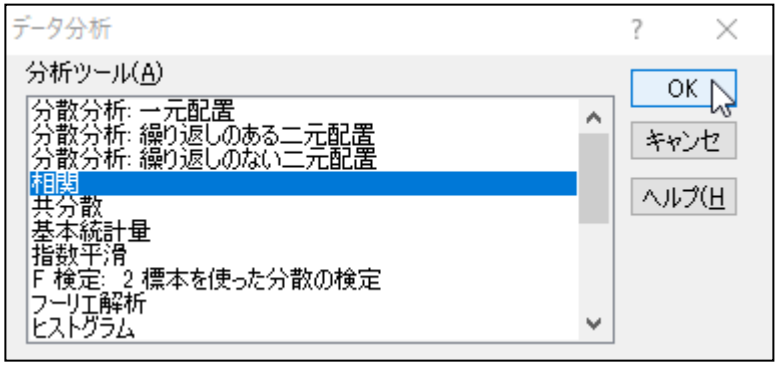
分析ツールによる相関係数の一括導出（相関行列）

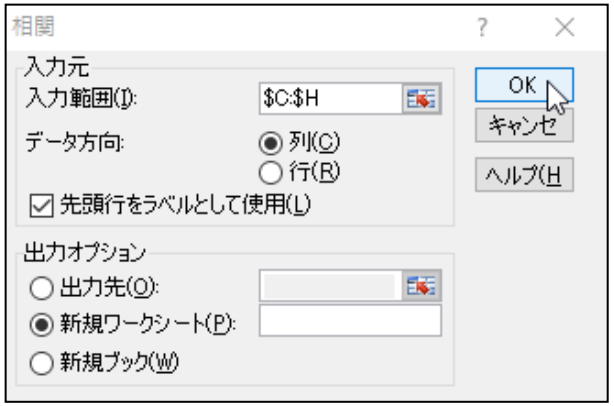
◆Excelの分析ツールから、様々な変数の組み合わせの相関係数を一括して導出できます。

- Excelの「分析ツール」の「相関」から、**相関係数の組み合わせを一括表示する相関行列**を作成できます。

Excel画面上側の「データ」タブの「データ分析」をクリックし、ダイヤログボックスから「相関」を選択し、「OK」をクリックします。
入力範囲としてシート〔1〕のC列の［東京：平均気温(℃)］～H列［浜松店：飲料販売量（本）］を選択します。
「先頭行をラベルとして使用」にチェックを入れ「新規ワークシート」を選択し、「OK」をクリックします。







- 出力された相関行列では、縦の列と横の行の交差する欄に各2変数の相関係数が記入されています。
  - Excel関数で導出した相関係数を確認できることに加え、シート〔1〕にある変数は全ての組み合わせで正の相関関係にあることが分かります。

分析ツールの「相関」による相関行列の出力

	東京: 平均気温(℃)	浜松: 平均気温(℃)	東京: 平均湿度(%)	浜松: 平均湿度(%)	東京店:飲料 販売量(本)	浜松店:飲料 販売量(本)
東京:平均気温(℃)	1					
浜松:平均気温(℃)	0.977	1				
東京:平均湿度(%)	0.493	0.589	1			
浜松:平均湿度(%)	0.525	0.561	0.809	1		
東京店:飲料販売量(本)	0.548	0.543	0.365	0.367	1	
浜松店:飲料販売量(本)	0.671	0.702	0.554	0.596	0.644	1

- 相関行列は、様々なパターンの2変数の組み合わせの直線的な関係を、一覧で確認したい場合に便利です。

散布図における単回帰分析

◆Excelの散布図においても、単回帰分析の結果を確認することができます。

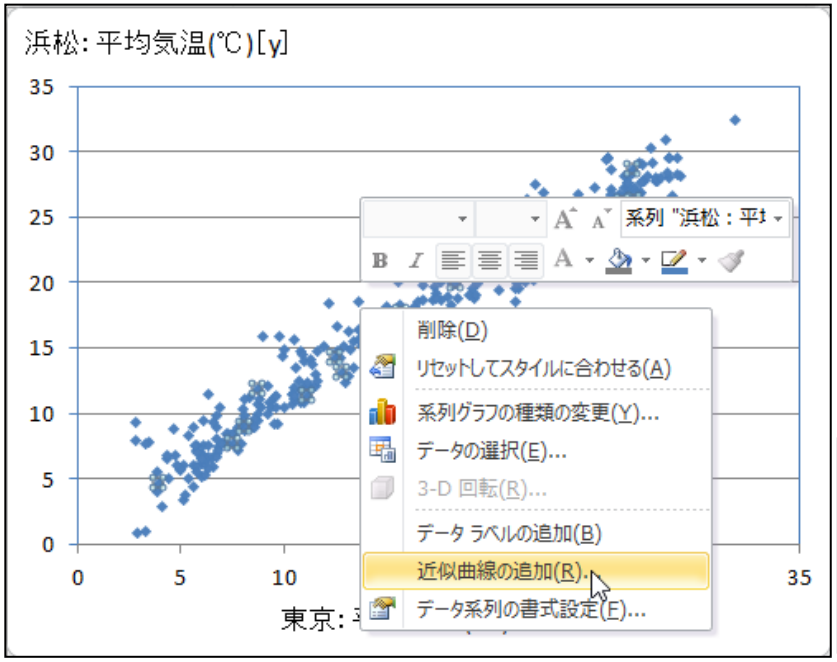
- Excelの散布図においても、説明変数が1種類の単回帰分析の結果を確認することができます。

  散布図を右クリックして表示されるメニューから「近似曲線の追加」をクリックすることで、二変数の関係を示す回帰線を引きます。
  表示されたダイアログボックスから「グラフに数式を表示する」「グラフにR-2乗値を表示する」にチェックを入れてください。

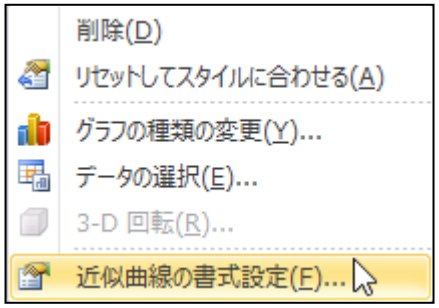
  - ダイアログボックスが表示されずに近似曲線が引かれた場合は、近似曲線を右クリックして「近似曲線の書式設定」から選択することができます。

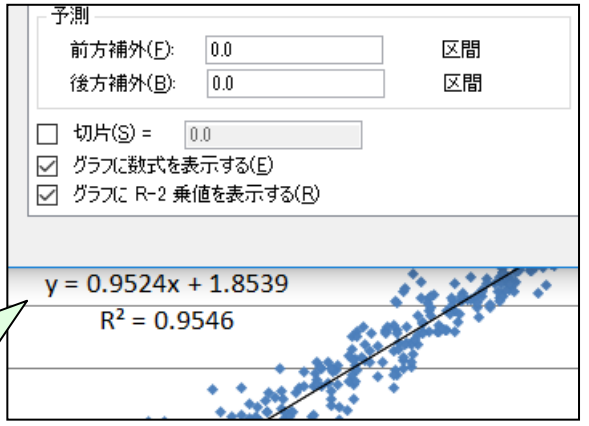
- 回帰分析では「変数xが1大きくなった場合に、変数yにどの程度の影響を与えるか？」を測定することができます。
  - 同時に表示されるR²は、決定係数と呼ばれる「0～1」の間をとる回帰分析の当てはまりの指標であり、後述します。
  - ここでは考察の対象外としますが、暗黙裡に「変数x［東京の気温(℃)］→変数y［浜松の気温(℃)］」の一方向の効果を仮定しています。

右クリックメニューから「近似曲線の追加」



「近似曲線の書式設定」から「グラフに数式を表示する」にチェック





散布図上に「東京の気温と浜松の気温の関係を表す数式」が表示されます。
この画面の結果では、
- **東京の気温が1℃増加したら、浜松の気温が0.9524℃増加する関係にあること**
- **東京の気温0℃に対応する浜松の気温は1.8539℃であること**

を示しています。

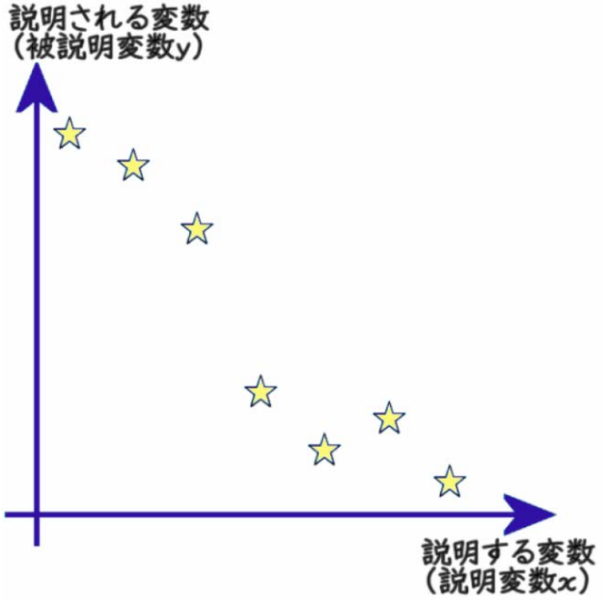
□ 続いて、回帰分析（最小二乗法）が「どのような発想で直線を引いているのか？」を紹介します。

回帰分析（最小二乗法）における回帰線導出のルール

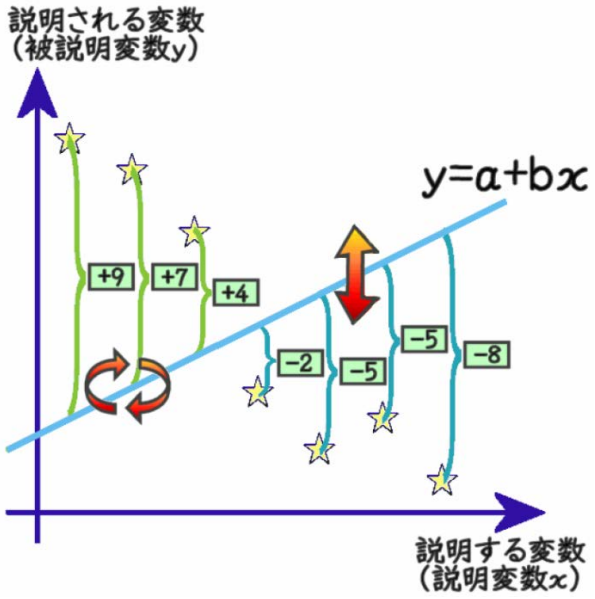
◆回帰分析は、差分を2乗の総和を最小化するルールで当てはまりの良い直線を引きます。

- 左下図のような説明される変数（被説明変数）yと説明する変数（説明変数）xの関係による点の星☆があり、**「2種類の変数の関係を示す当てはまりの良い直線（y=a+bx）を引く」ためのルール**を考えます。
  - aの値を変更することで直線の高さを変えたり、bの値を変更することで直線の傾きを変えて、当てはまりの良い直線を引くことを考えます。
- 「星と直線の差分（距離）の総和を最小化するルール」が良さそうに思いますが、正と負の差分が相殺してしまいます。
  - 点のはるか上に線を引くことで、絶対値が大きい負の差分を得ることができ、差分の総和を負の値にすることも可能です。
- 「星と直線の差分を二乗して、全て正の値にしてから総和を最小化するルール」なら、負の差分の問題がありません。
- 回帰分析は、**「星と直線の差分を2乗してから総和を最小化するルール」**で**直線（⇒回帰線）**を引きます。

☆に表される「xとyの関係を示すための当てはまりの良い直線」を引くためのルールを考えたい。



「☆と直線の差分の総和を最小化」するルールでは、正と負の差分で相殺し、当てはまりの良い直線にならない。



☆と直線の差分を2乗してから、「『全て正の値に変換した差分』の総和を最小化するルール」なら、当てはまりの良い直線が引ける。

2乗した差分の総和を最小化するため、**回帰分析は最小二乗法**とも呼ばれます。

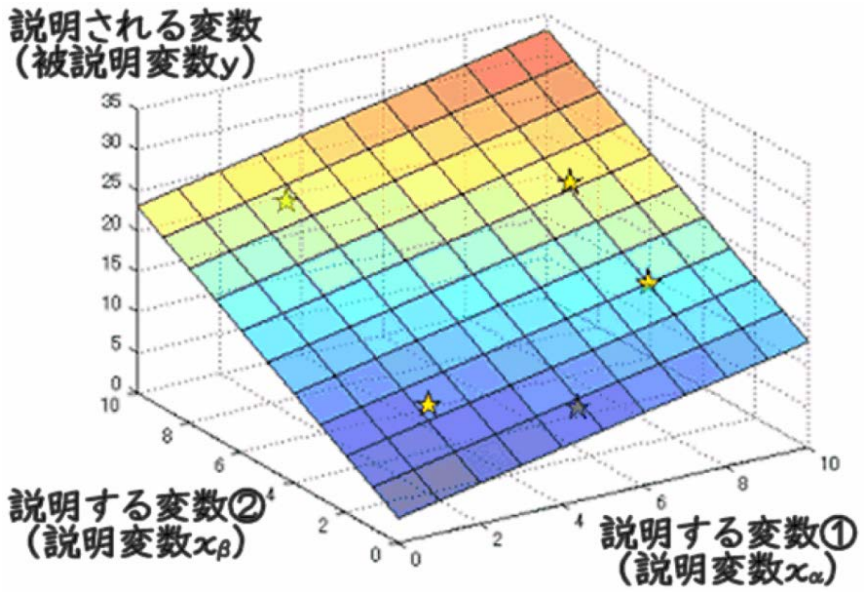
- 最小二乗法の方が分析における利用頻度は高いですが、「星と直線の差分の絶対値を最小化するルール」の最小絶対値法という手法もあります。

重回帰分析（説明変数が2種類以上の回帰分析）

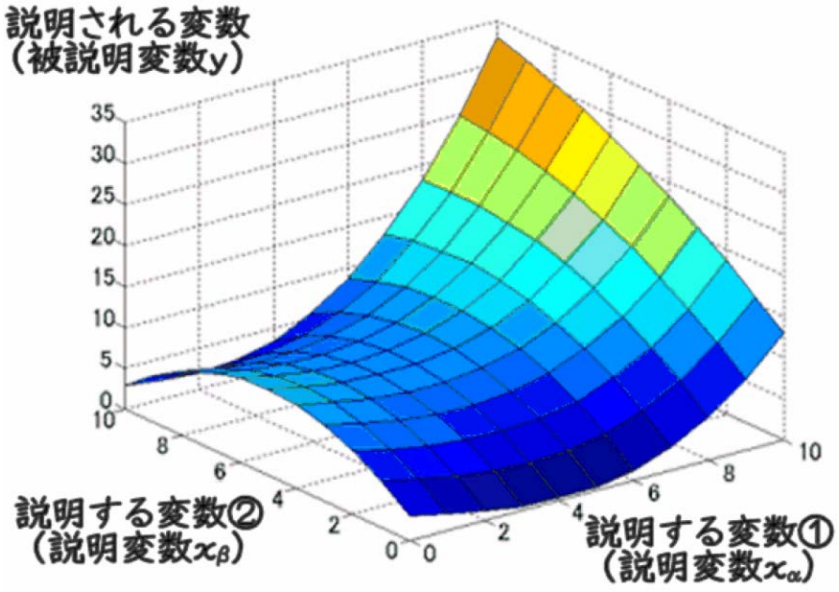
◆2種類以上の説明変数による回帰分析のことを重回帰分析と呼びます。

- Excelの分析ツールを利用すれば、**「説明変数が2種類以上の重回帰分析」**を行えます。
  - Excelの散布図における「近似曲線」では単回帰分析しかできませんが、Excelの分析ツールでは説明変数16種類までの重回帰分析を実行できます。
- 単回帰分析においては、最小化の対象は「星と直線の差分の二乗和」でしたが、重回帰分析においては当てはめる直線（⇒回帰線）が左下図のように板（⇒回帰平面）になります。
  - 説明変数が3種類以上で四次元以上になると想像しにくいですが、考え方は二次元から三次元に増える場合と同じです。
  - 回帰平面は各説明変数の平均値と被説明変数の平均値を通り、回帰平面の傾きは各説明変数から被説明変数への平均的な効果に対応します。
- 標準的な回帰分析では星と直線・平面との差分を考えますが、（直線・平面を前提としても）説明変数をあらかじめ2乗するなどの変換しておき、説明変数を元の尺度に戻せば、右下図のように曲線の関係を表すこともできます。
  - 「xの2乗」を説明変数に入れると、xが1、2、3と増えた場合に「xの2乗」は1、3、9と増加し、もとのxの尺度では曲線（曲面）の回帰線となります。

重回帰分析のイメージ



説明変数を変換して曲線の関係も分析可能



回帰分析用のデータセット（ダミー変数の作成）

◆土日祝を表すダミー変数を追加して、回帰分析用のデータセットを作成します。

- ［東京店：飲料販売量(本)］を被説明変数として［東京：平均気温(℃)］と［東京：平均湿度(%)］を説明変数として、重回帰分析を行うことを考えます。
  - シート〔3〕の相関行列で示したようにシート〔1〕の全ての変数は正の相関関係にあります。相関関係としては、気温または湿度が高まれば、飲料販売量が増える直線的な関係にあります。しかし、相関係数や単回帰分析では気温か湿度のどちらかだけの効果か、両方ともに効果があるのかを判別できません。

💻 シート〔1〕から［日付］［曜日（休日）］［東京：平均気温(℃)］［東京：平均湿度(%)］の各変数をコピーして、新規シートのA列～D列へ貼りつけます。

- 飲料販売量に影響を与える説明変数として、「平日」か「土日祝」の違いをダミー変数として追加することを考えます。

💻 E2のセルに『=IF(OR(B2="月",B2="火",B2="水",B2="木",B2="金"),0,1)』と入力すると、1が表示されます。
  - Excel関数IFの中にExcel関数ORを入れ込み、B2の値が「月」「火」「水」「木」「金」のいずれかと一致すれば0を出力し、そうでなければ1を出力します。

💻 表示後にE2のセルの右下の■をダブルクリックして、0か1の表示を最下段まで引き延ばし、変数名に「土日祝ダミー」と入力します。

- 作成した［土日祝ダミー］は、［曜日（休日）］に基づき、平日であれば0、土日祝であれば1の値を持ちます。
- 本講座の冒頭で「回帰分析は連続的な値（間隔尺度や比率尺度）をとる変数間の分析」と紹介しましたが、説明変数に関しては、名義尺度や順序尺度も**「0か1の値をとるダミー変数」**として、回帰分析に利用できます。
  - より高度な回帰分析には、被説明変数を名義尺度や順序尺度とするものもありますが、標準的な回帰分析の範疇外となります。

💻 シート〔1〕から［東京店：飲料販売量(本)］をコピーして、新規シートのF列へ貼りつけます。

作成した回帰分析用のデータセット

	A	B	C	D	E	F
1	日付	曜日（休日）	東京：平均気温(℃)	東京：平均湿度(%)	土日祝ダミー	東京店：飲料販売量(本)
2	2016/1/1	金(休)	7.5	45	1	296
3	2016/1/2	土(休)	7.3	51	1	414
4	2016/1/3	日(休)	9.3	61	1	343
5	2016/1/4	月	9.2	60	0	514
6	2016/1/5	火	10.9	51	0	429

Excelの「分析ツール」の回帰分析における指定

◆回帰分析のダイアログボックスに、被説明変数と説明変数を指定します。

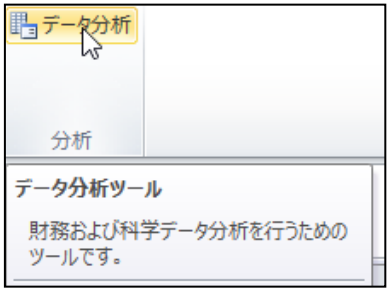
- Excelの「データ分析」のダイアログボックスにおいて、「回帰分析」を指定し、入出力の変数と出力先を指定します。

Excel画面上側の「データ」タブの［データ分析］をクリックし、ダイアログボックスから「回帰分析」を選択し、「OK」をクリックします。

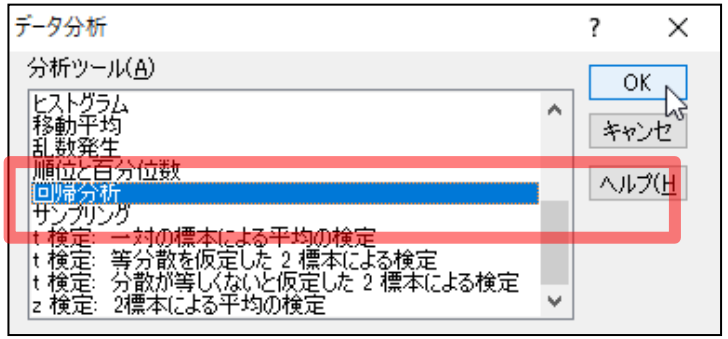
入力Y範囲には、「説明される変数（1列のみ）」として「F列の範囲」を指定し、入力X範囲には「説明する変数（複数列でも可）」として、「C列～E列」を指定します。ボックス内の「ラベル」にチェックを入れて「OK」をクリックします。

- 回帰分析の結果を異なるExcelシートに出力したい場合は初期設定の「新規ワークシート」にチェックを入れて「OK」をクリックします。一方で、分析用のデータセットと同じシート内に表示したい場合は、［一覧の出力先］に分析結果を表示したい範囲の左上のセルを指定します。

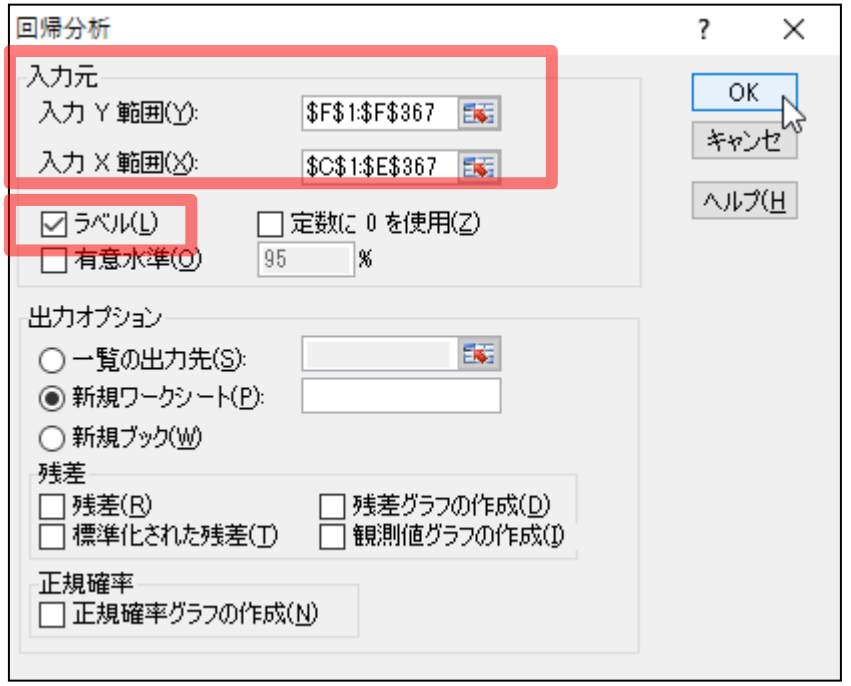
「データ分析」をクリック



「回帰分析」を選択して「OK」をクリック



「入力Y範囲」「入力X範囲」を記入して「OK」をクリック



Excelの回帰分析における結果出力

◆Excelの分析結果は、「回帰分析全体」「個別の説明変数」の2種に分かれて表示されます

- Excelの回帰分析の結果出力は3ブロックに分かれ、**上側の2つのブロックが［回帰分析全体に関する出力］**に相当し、**最下段のブロックが［個別の説明変数に関する出力］**の結果になります。

Excelの回帰分析の結果出力

概要

回帰統計	
重相関 R	0.720
重決定 R2	0.518
補正 R2	0.514
標準誤差	62.689
観測数	366

分散分析表

	自由度	変動	分散	観測された分散比	有意 F
回帰	3	1530902	510300.6	129.849	4.16E-57
残差	362	1422641	3929.946		
合計	365	2953542			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	374.963	14.569	25.738	8.80898E-84	346.314	403.613	346.314	403.613
東京:平均気温(℃)	5.880	0.491	11.965	4.98243E-28	4.914	6.847	4.914	6.847
東京:平均湿度(%)	0.490	0.232	2.115	0.035070468	0.035	0.946	0.035	0.946
土日祝ダミー	-86.739	6.968	-12.449	7.40285E-30	-100.441	-73.037	-100.441	-73.037

回帰分析全体に関する出力

個別の説明変数に関する出力

結果出力の上段

結果出力の中段

結果出力の下段

□ この講座では回帰分析の結果出力の上段と中段を簡潔に説明し、予測等に利用する下段を重点的に説明します。

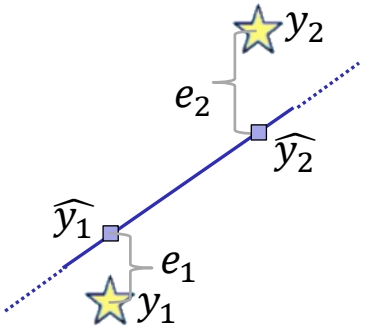
出力の上段における決定係数〈重決定 R2〉

◆上段出力の2行目の決定係数〈重決定 R2〉は、回帰分析の当てはまりの指標です。

- 上段出力の2行目の**決定係数**〈重決定 R2〉は、**0から1の値をとる回帰分析の当てはまりの指標**です。
  - Excelの回帰分析の出力にある〈重決定 R2〉〈重相関 R〉〈補正 R2〉という用語は統計学においても、実際の分析レポートにおいても、ほとんど使われず、Excelの分析ツールの独自用語です。本教材では、Excelの独自用語は〈〉で括って示し、統計学で一般的に使われる用語の後に記載しています。
  - 散布図における単回帰分析でもR2として表示され、重要度の高い2行目の決定係数〈重決定 R2〉から説明した後、1行目の〈重相関 R〉を説明します。

- 回帰分析では被説明変数を変数 y で表す時、回帰線による**予測値**を$\hat{y}$で表し、**推定エラー**を eで表し、i 番目の標本に関して $y_i = \hat{y}_i + e_i$ と書けます。
  - 予測値の $\hat{y}$ は「ワイハット」と読み、推定エラー（error）の頭文字に由来する e は「イー」と読みます。
  - 予測値は当てはめ値と呼ばれることもあり、推定エラーは残差と呼ばれることもあります。

予測値と推定エラーのイメージ



- 予測値 $\hat{y}$ の平均値は、被説明変数 y の平均値 $\bar{y}$ に一致する性質があり、
  $\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 + \sum_{i=1}^{n}e^2$ という式が必ず成り立ちます。
  - この式は「被説明変数の偏差平方和＝予測値の偏差平方和＋推定エラーの二乗和」を意味しています。
  - 推定エラー e の平均値は0という性質があり、「推定エラーの二乗和」は「推定エラーの偏差平方和」でもあります。

- この式の両辺を左辺で割って移項すると、**決定係数**〈重決定 R2〉$= \dfrac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} = 1 - \dfrac{\sum_{i=1}^{n}e^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$ となり、

決定係数〈重決定 R2〉は $\dfrac{\text{予測値の偏差平方和}}{\text{被説明変数の偏差平方和}} = 1 - \dfrac{\text{推定エラーの二乗和}}{\text{被説明変数の偏差平方和}}$ を意味しています。

  - 上記の式の各項は、二乗した値の総和をとっているので0以上の値となり、決定係数は0以上1以下の値になります。

- **決定係数**〈重決定 R2〉**は、被説明変数の偏差平方和に占める予測値の偏差平方和**を示し、**被説明変数の変動を予測値の変動で説明できる割合に相当**します。

回帰分析の結果出力の上段

◆結果出力の上段は、回帰分析全体に関する出力を示しています。

- 出力2行目の**決定係数**〈重決定 R2〉は、前スライドで示した**回帰分析の当てはまりの指標**です。
  - 今回の出力では、予測値の変動によって被説明変数の変動の51.8%を説明できていることを示しています。

結果出力の上段

概要	
回帰統計	
重相関 R	0.720
重決定 R2	0.518
補正 R2	0.514
標準誤差	62.689
観測数	366

- 出力1行目の〈重相関 R〉は、2行目の決定係数〈重決定 R2〉の正の平方根です。
  - 〈重相関 R〉は、Excelの独自用語であることに加えて、2行目の決定係数〈重決定 R2〉から算出できることもあって、実際の分析資料において〈重相関 R〉を表示することは、ほとんどありません。

- 出力3行目の**自由度調整済み決定係数**〈補正 R2〉は、**自由度（＝標本数-説明変数の種類数）を考慮した回帰分析の当てはまりの指標**です。

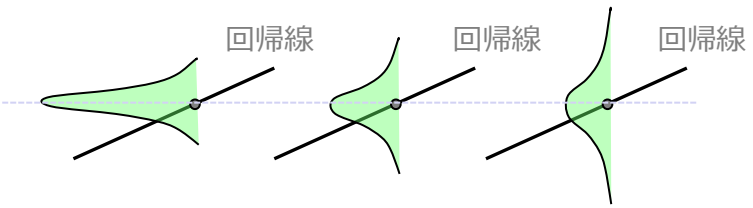
$$\text{自由度調整済み決定係数〈補正 R2〉} = 1 - (1 - \text{決定係数}) \cdot \frac{(\text{標本数} - 1)}{(\text{標本数} - \text{説明変数の種類数})}$$

（標本数 – 説明変数の種類数）：自由度

  - 決定係数は前スライドに示したように被説明変数の変動に占める（全ての説明変数での）予測値によって説明できる割合を示しています。無関係な説明変数であっても、説明変数の種類を増やせば、予測値で説明できる割合に相当する決定係数は必ず上昇します。このため、説明変数を増やすことによって低下する要素として、上記の式の分数箇所を加えた当てはまりの指標が「自由度調整済み決定係数」です。
  - 標本数が膨大であれば、上記の式の分数箇所は1に接近し、決定係数〈重決定 R2〉と自由度調整済み決定係数〈補正 R2〉は、僅差となります。
  - 自由度の考え方は、講座3-2で示した「（母集団として）の分散と標本分散」の関係と同一であり、標本数と同じだけ説明変数の種類があれば、全ての標本を説明できて当たり前なので「説明できない可能性がある余地」に相当する自由度（＝標本数-説明変数の種類数）で調整しています。

- 出力4行目の**〔推定エラーの〕標準誤差**は、**回帰式の推定エラー（e）の平均的なバラツキ**を示しています。
  - 回帰分析の出力下段にも「標準誤差」と同様の表記があるなど、Excelの出力において、語句の混同を招く部分、言葉足らずの部分は本教材において〔〕で補足します。
  - 決定係数の導出にも利用した「推定エラーの二乗和」を自由度で割ると、〔推定エラーの〕不偏分散が導出でき、その正の平方根をとることで〔推定エラーの〕標準誤差となります。
  - 変数の分布に関するバラツキを標準**偏**差、推定値に関するバラツキを標準**誤**差といいます。

右側ほど〔推定エラーの〕標準誤差が大きい



回帰分析の結果出力の中段（帰無仮説の考え方）

◆結果出力の中段は、回帰分析の全ての変数が無意味な場合の検定結果を示しています。

- 回帰分析の結果の中段にある〈分散分析表〉は、『**切片以外の全ての説明変数は無効⇒切片以外の説明変数の真の係数は全て0である**』という**帰無仮説の検定**を行っています。
  - 利用可能な標本を用いて変数間の関係を推測する場合、誤差等の影響によって考察すべき母集団の変数間の関係を正しく把握できません。「説明変数は無効」「真の係数はゼロ」とは、考察すべき母集団において「変数間が無関係であること」を指しています。

帰無仮説の考え方

統計学では『効果がない（無効）』や『異なっていない（同質）』を主張する仮説を**帰無仮説(きむかせつ)**といいます。一方、帰無仮説の反対側の『効果がある（有効）』や『異なっている（異質）』を主張する仮説を**対立仮説**といいます。対立仮説を直接肯定することが難しいため、帰無仮説を否定することで**間接的に対立仮説を肯定するという手続き**をとります。なお、統計学の用語では「特定の判定基準のもとで帰無仮説を否定する」ことを「**帰無仮説を棄却（ききゃく）する**」や「**統計的に有意（ゆうい）な効果/違いがある**」といいます。

- 中段の一番右側にある〈有意F〉は『**切片以外の説明変数の真の係数は全て0である**』という帰無仮説のもとで、偶然の誤差の影響によって標本の関係が観測されてしまう確率の上限を示しています。
  - 〈有意 F〉はExcelの独自用語です。統計学では、この指標の利用頻度は少ないものの『複数制約のP値』と呼ぶ方が一般的です。

結果出力の中段

分散分析表					
	自由度	変動	分散	観測された分散比	有意 F
回帰	3	1530902	510300.6	129.849	4.16E-57
残差	362	1422641	3929.946		
合計	365	2953542			

「4.16E-57」は「4.16*0.1^57」を表し、0が57個並ぶほど、限りなく0に近い数値です。
『全ての変数が無効』であることを前提とすれば、極めて不自然なことが起こったことを意味しています。

- 今回の出力のように〈有意F〉が十分にゼロに近ければ、「『**全ての変数が無効』という帰無仮説を前提とすれば、確率的に極めて珍しいことが起こった**、つまりは『**前提としていた帰無仮説は不自然で、ほぼ確実に効果のある説明変数がある。**』」という解釈になります。
  - 統計学は確定的な結論が出る学問ではありませんが、ある仮定を前提にして「無理がある/不自然である」ことを示して、前提としていた仮定を吟味する帰無仮説の検定の考え方は、数学における背理法と類似しています。

〔推定〕係数

◆〔推定〕係数は「説明変数1単位の増加→被説明変数への効果」の推定値を指しています。

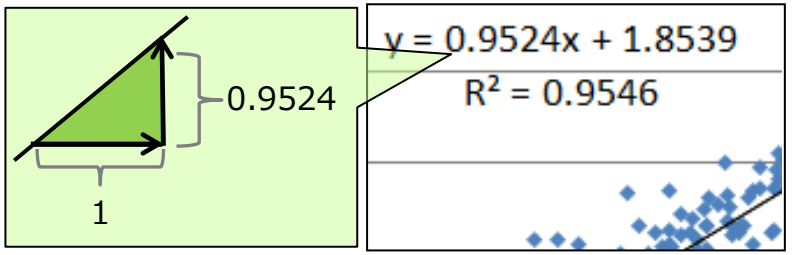
- **切片の〔推定〕係数**は、他の**全ての説明変数がゼロの場合における回帰分析による予測値**に該当します。
  - 今回の出力では、仮想的に気温0℃、湿度0%の状態を想定した場合、平日の飲料販売量の予測値が約375本であることを示しています。

結果出力の下段（〔推定〕係数）

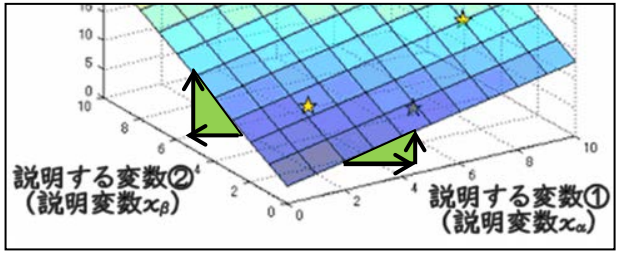
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	374.963	14.569	25.738	8.80898E-84	346.314	403.613	346.314	403.613
東京: 平均気温(℃)	5.880	0.491	11.965	4.98243E-28	4.914	6.847	4.914	6.847
東京: 平均湿度(%)	0.490	0.232	2.115	0.035070468	0.035	0.946	0.035	0.946
土日祝ダミー	-86.739	6.968	-12.449	7.40285E-30	-100.441	-73.037	-100.441	-73.037

- **切片以外の〔推定〕係数**は、**説明変数の1単位の増加→被説明変数への効果の推定値**を示しています。
  - 今回の出力では、気温が1℃上がれば飲料販売量は約6本増え、湿度が1%上がれば、飲料販売量が0.5本増えることを示しています。
  - 効果の方向の仮定を誤って逆にすると、「飲料販売量が上がれば、気温が上がる」ので「涼しくするために飲み物を控えよう」という誤った結論になります。
  - 説明変数の単位を1℃単位から10℃単位にするなど、入力値を0.1倍にした場合、推定係数は調整されて10倍になります。
- **ダミー変数に関する〔推定〕係数**は、**ダミー変数が1となる場合に予測値が変化する水準**を示しています。
  - 今回の出力では、土日祝ダミーの推定値は「-87」であり、平日を基準として土日祝の回帰線は、87本少なくなるように平行移動します。
  - 今回のダミー変数は、利用頻度の高い「定数項ダミー」という回帰線が平行移動するタイプで設定しましたが、「係数ダミー」と呼ばれる他の説明変数の推定値（傾き）を変えるタイプもあります。［例：平日と土日祝では「気温→飲料販売量」の効果の大きさ（傾き）が異なる想定］

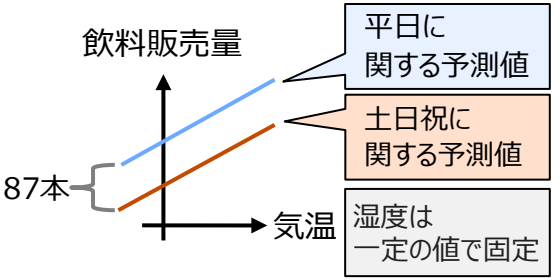
単回帰分析における推定係数（傾き）



重回帰分析における推定係数（傾き）



ダミー変数の出力結果イメージ



〔推定係数の〕標準誤差と〔信頼区間の〕下限・上限

シート〔5〕回帰分析の出力
出力下段の〔推定係数の〕標準誤差

◆〔推定係数の〕標準誤差と〔信頼区間の〕下限・上限は、推定係数の不確かさに関する指標です。

- 〔推定〕係数は利用可能な標本から推定するため、母集団において真の関係を表す係数からの乖離が見込まれます。
  - 真の係数とは、無限個の標本に対応する母集団（考察の対象とする全ての標本）を利用した場合に確認できる正しい関係を指しています。
  - 乱数に基づくシミュレーションにおいては真の係数を設定できますが、実証分析における真の係数は分析者にとって最後まで分からない数値です。
- Excel出力の**〔推定係数の〕標準誤差**は、**「推定係数の不確かさ（⇒推定精度の悪さ）」**に該当します。
  - 説明変数の単位を1℃単位から10℃単位にするなど、入力値を0.1倍にした場合、〔推定係数の〕標準誤差は調整されて10倍になります。
- **〔信頼区間の〕下限および上限の95%は、信頼係数95%で真の係数がありそうな範囲**を示しています。
  - 回帰分析実行時のダイアログボックスで90%や99%の信頼係数を入力して指定することができ、結果出力における右側の2列の値が変化します。
  - 〔推定〕係数によって一点で示されている値が、真の係数に一致すると考えるのは楽観的ですが、真の係数は推定係数を中心にその付近に存在すると考えるのが自然です。

信頼係数

ここで**確率**という言葉を使わず、**信頼係数**という言葉を用いるのは**「真の係数は分析者にはっきりと分からないだけで、確定した値がある」**という考え方に由来します。例えば、地震の研究者への『3世紀の日本で震度7以上の大地震が起こったか？』という問いに対して、「私は十中八九、起こったと考える」とは答えられても、「確率80%～90%で起こった」とは答えません。ただし、信頼係数という言葉に馴染めなければ、**信頼係数は「（利用可能な標本に基づく分析者の）主観的な確率」**と読みかえてもかまいません。

- 〔信頼区間の〕下限および上限 95%の値は、おおよそ推定係数+(±2)×〔推定係数の〕標準誤差となります。
  - 今回の出力では、気温の推定係数の信頼区間95%の上限の概数は5.88+2×0.491=6.862（Excel出力：6.847）であり、下限の概数は5.88-2×0.491=4.898（Excel出力：4.914）であり、それぞれ近似値になっています。

結果出力の下段（〔推定係数の〕標準誤差、〔信頼区間の〕下限・上限）

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	374.963	14.569	25.738	8.80898E-84	346.314	403.613	346.314	403.613
東京: 平均気温(℃)	5.880	0.491	11.965	4.98243E-28	4.914	6.847	4.914	6.847
東京: 平均湿度(%)	0.490	0.232	2.115	0.035070468	0.035	0.946	0.035	0.946
土日祝ダミー	-86.739	6.968	-12.449	7.40285E-30	-100.441	-73.037	-100.441	-73.037

【参考】乱数シミュレーションに基づく回帰分析

◆推定係数の標準誤差は、変動する要因を乱数シミュレーション等で確認します。

- 推定係数の標準誤差は、下の表内の要因によって減少します。

推定係数の標準誤差を低下させる要因と確認方法

項目	〔推定係数の〕標準誤差を低下させる要因	〔推定係数の〕標準誤差の増減との関係の説明	〔推定係数の〕標準誤差が低下することの確認方法
[1]	標本数の増加	標本数が多ければ、推定精度は高まるため、推定係数の標準誤差は小さくなります。	シート〔4〕のデータセットを縦に並べて2年分（732標本）の扱いで回帰分析を実行し、比較してください。
[2]	各説明変数のバラツキ（標準偏差）の増加	各説明変数のばらつきが大きい方が、推定係数を測りやすく、推定係数の標準誤差は小さくなります。	シート〔参2〕における説明変数の標準偏差を指定するF4およびF5のセルの値を増加させてください。
[3]	説明変数同士の相関係数の絶対値の低下	説明変数同士が相関が小さいと、どの説明変数の効果かを判別しやすく、推定係数の標準誤差は小さくなります。	シート〔参2〕における説明変数間の相関係数を指定するB6のセルの値を0に近づけてください。
[4]	回帰式全体に関するエラーのバラツキ（標準偏差）の低下	回帰式全体に関するエラーのバラツキが小さいと、推定係数を測りやすく、推定係数の標準誤差は小さくなります。	シート〔参2〕における回帰式のエラーの標準偏差を指定するB7のセルの値を減少させてください。

- 項目[1]に関して、膨大な標本数（→標本数の多いビッグデータ）では、推定係数の標準誤差は0に接近し、ほとんど全ての帰無仮説を棄却できます。
- 項目[3]に関して、説明変数同士の相関係数が±1（→完全な多重共線性）の場合、どちらの説明変数の効果かを判別できず、分析できなくなります。

◘ シート〔参2〕では、シミュレーションに基づいて回帰分析の出力の挙動を確認できます。

キーボードの［F9］または「数式」タブの「再計算実行」をクリックして、出力の挙動を確認してください。

シート〔参2〕乱数シミュレーションに基づく回帰分析

水色のセルを変化させて、設定を変更できます。キーボードの[F9]ボタンまたは「数式」タブの「再計算実行」で乱数が再発生し、回帰分析の出力が再計算されます。						
切片	15	標本数	50			
X_αの係数	3	X_αの平均	10	X_αの標準偏差	7	←項目[2]の確認は、これらの値を増加させる[範囲：0～]
X_βの係数	5	X_βの平均	20	X_βの標準偏差	3	
X_αとX_βの相関係数	0.3	←項目[3]の確認は、この値を0に近づける[範囲：-1～1]				説明変数x_αとx_βは二次元正規分布にて発生
回帰式のエラーの標準偏差	9	←項目[4]の確認は、この値を減少させる[範囲：0～]				回帰式のエラーは、正規分布にて発生
真のデータ発生プロセス：	Y=15+3Xα+5Xβ+N(0, 9^2)					
回帰分析の出力：	Y=15.15+3.41Xα+4.77Xβ+N(0, 7.1^2)					↓Excelの分析ツール風の出力

t〔値〕

◆t〔値〕は、基準精度で評価した推定係数に該当します。

$$t\text{値} = \frac{\text{推定係数}}{\text{推定係数の標準誤差}}$$

- **t値**は、推定係数の標準誤差で割ることで**基準精度で評価した推定係数**です。
  - 今回の出力においては、いずれの説明変数のt値も絶対値で2を超えており、説明変数のそれぞれに被説明変数への効果があったと見なせます。
  - 説明変数の単位を1℃単位から10℃単位にするなど、入力値を0.1倍にした場合、分子と分母のそれぞれが10倍されてt値は変わりません。

結果出力の下段（t〔値〕）

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	374.963	14.569	25.738	8.80898E-84	346.314	403.613	346.314	403.613
東京: 平均気温(℃)	5.880	0.491	11.965	4.98243E-28	4.914	6.847	4.914	6.847
東京: 平均湿度(%)	0.490	0.232	2.115	0.035070468	0.035	0.946	0.035	0.946
土日祝ダミー	-86.739	6.968	-12.449	7.40285E-30	-100.441	-73.037	-100.441	-73.037

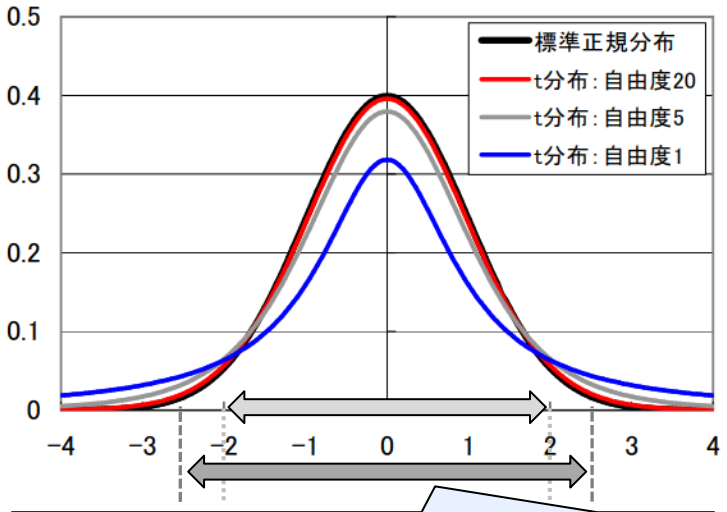
- **t値を用いて『説明変数の被説明変数への真の効果が0である』という帰無仮説を検定**することができます。

t分布と標準正規分布

説明変数の「真の効果がゼロ」という帰無仮説が正しければ、**t値はt分布**に従うことが知られています。t分布は、右図のように0を中心に左右対称にばらつく山形の分布です。なお、**t分布は自由度（標本数-説明変数の種類数）が上がると、標準正規分布に接近します。**自由度が20以上であれば、実務上は標準正規分布と考えて差し支えありません。

自由度の高いt分布や標準正規分布において、絶対値で2からはみ出す面積の割合は、プラス側とマイナス側のそれぞれ約2.5%で両側約5%です。

t分布と標準正規分布



帰無仮説を前提とすれば、
両側5%基準：95%が±2.0の範囲に収まる ⇔
両側1%基準：99%が±2.6の範囲に収まる ⇔

実際の分析で利用頻度の高い「両側5%の基準での検定」の考え方

t値の絶対値が2未満	⇒帰無仮説の前提で5%以上は発生する自然なt値 **⇒帰無仮説が誤っているとは考えにくい。**
t値の絶対値が2以上	⇒帰無仮説の前提で5%以下でしか発生しない不自然なt値 **⇒帰無仮説が誤っていると考え、効果のある変数と見なせる。**

P-値

◆P-値は、統計的検定の境目となる確率を表しています。

- 前スライドに示したt値は、次のような場合に利用しにくい指標となります。
  - ◆ 両側5%以外の基準で検定を行う場合　（⇒ t値の絶対値の2を基準とする目安が利用できない）
  - ◆ 自由度（標本数-説明変数の種類数）が20未満などの場合（⇒ t分布の標準正規分布への接近が不十分）
- こうした場合に便利で、**自由度を考慮した統計的検定の境目を出力しているのがP-値**です。

（自由度の高い）回帰分析において、t値が2以上2.6未満の範囲に入った場合は「両側5%基準では帰無仮説が棄却できる（⇒両側5%で有意）」の一方で「両側1%基準では帰無仮説が棄却できない（⇒両側1%では有意ではない）」ことになります。

⇨ 1%と5%の間に境目となる基準の値があったはずで、P-値はこの境目となる値を導出しています。

- 今回の出力では、湿度のt値は2.115であり、5%と1%の間の検定の境目があるはずです。右隣のP-値を見ると、それが3.5%であることが分かります。

結果出力の下段（P-値）

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	374.963	14.569	25.738	8.80898E-84	346.314	403.613	346.314	403.613
東京: 平均気温(℃)	5.880	0.491	11.965	4.98243E-28	4.914	6.847	4.914	6.847
東京: 平均湿度(%)	0.490	0.232	2.115	0.035070468	0.035	0.946	0.035	0.946
土日祝ダミー	-86.739	6.968	-12.449	7.40285E-30	-100.441	-73.037	-100.441	-73.037

- 両側5%の判定基準に関しても、「t値が絶対値で2を超えているか？」の目安よりも「P−値が0.05を切っているか？」の方が、自由度を考慮していて正確です。

統計的検定と説明変数の効果

t値やP-値を用いた検定は『推定精度をふまえて、効果が0なのか？』を基準に判定しています。このため、「『推定係数の絶対値が小さくとも、推定精度が高い変数』を効果がありそうな説明変数」と判定します。一方で「『推定係数の絶対値が大きくとも、推定精度が低い変数』を効果がなさそうな説明変数」と判定します。**t値やP-値に基づく効果のある変数の表示は、良くも悪くも不確かさを考慮しているので、推定係数の（絶対値の）大きさに基づく表示や利用と使い分ける必要**があります。

分析結果の考察と活用方法

◆ 回帰分析の分析結果は、説明変数の効果測定や予測に活用できます。

- 今回の出力では**決定係数**〈重決定 R2〉が0.518であり、予測値の変動によって、被説明変数の変動の51.8%の説明ができています。
  - 決定係数の評価に絶対的な水準はないものの、回帰分析の設定に改善の余地が大きい決定係数の水準であるともいえます。
  - 説明変数を追加すれば決定係数は必ず上がりますが、分析目的によっては低い決定係数でも問題のない場合もあります。
- 回帰分析の式の形を変更したり、説明変数を追加したりして、**回帰分析の設定を改善することが可能**です。
  - 今回の回帰分析の設定では、気温が「0℃→1℃」でも「30℃→31℃」でも、飲料販売量に同一の効果があると仮定していました。「気温の2乗」や「気温の3乗」を説明変数に追加することで、気温の水準によって飲料販売量への効果が異なる可能性を考慮することができます。
  - 今回の回帰分析の設定では、気温と湿度は独立して、飲料販売量に効果があることを仮定していました。「気温×湿度」を説明変数に追加することで、気温と湿度の相乗効果の可能性を考慮することができます。
  - 今回の回帰分析の設定では、「土日祝ダミー」によって平日と土日祝では飲料販売量の水準が異なる可能性を考慮しましたが、土・日・祝のそれぞれで飲料販売量の水準が異なる可能性があります。土・日・祝のそれぞれに異なるダミー変数を設定することもできます。
  - 実証分析において、回帰分析の式の形はどうやっても近似にすぎませんが、現実を表すために「無理のない近似」となっている必要があります。
- **回帰分析では、不確かさを踏まえて説明変数の被説明変数への効果を導出し、予測に利用**することができます。

回帰分析の活用例

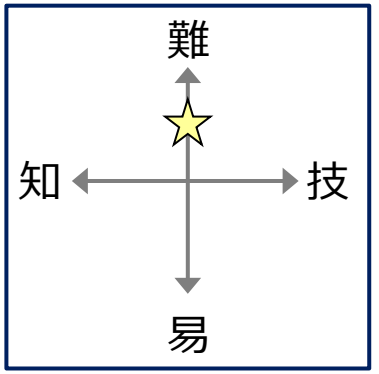
◆「予想温度▲℃、予想湿度■%の平日における飲料販売量の予測」など、与えられた設定における予測が可能となる。

⇨ **与えられた設定に基づく予測販売量を導出することで、最適な仕入れ量を検討することができる。**

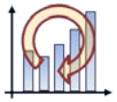
◆ 複数の変数の関係を調整し、不確かさを考慮した上で、説明変数が被説明変数に与える効果を導出できる。

⇨ **説明変数の効果の有無に関する水掛け論を防ぎ、統計的検定を踏まえて定量的・客観的な基準で評価・議論できる。**

総務省　ICTスキル総合習得教材



［コース3］データ分析



3-5：人工知能と機械学習

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_3_5.pdf

	1	2	3	4	5
［コース1］データ収集					
［コース2］データ蓄積					
［コース3］データ分析					●
［コース4］データ利活用					

本講座の学習内容［3-5：人工知能と機械学習］

【講座概要】

- 人工知能（AI)の種類と人工知能の概念を紹介します。
- 正解に相当する「教師データ」の状況に応じた機械学習の分類を説明します。
- 「回帰分析」「決定木」「k平均法」などの統計的機械学習の分析手法と用途を示します。
- 機械学習において、特に注目を集めているニューラルネットワークとディープラーニングの特性と基本的な用途を説明します。

【講座構成】

座学

- [1] 人工知能のイメージと定義
- [2] 機械学習の分類
- [3] 統計的機械学習の分析手法
- [4] ニューラルネットワークとディープラーニング

【学習のゴール】

- ✓ 人工知能（AI）の分類と定義を理解する。
- ✓ 正解に相当する「教師データ」の状況に応じた機械学習の分類を理解する。
- ✓ 統計的機械学習の分析の用途を把握する。
- ✓ ニューラルネットワークとディープラーニングの特性と基本的な用途を理解する。

人工知能の大分類

◆人工知能（AI）は大きく、汎用人工知能と特化型人工知能に分類されます。

- 本コースのこれまでの講座では、人間が行うデータクレンジング、データの可視化、データ分析を紹介してきました。
- 一方、データ分析をはじめとするデータの利活用を「人工知能に代行させる・任せる」といった形で**人工知能（AI：Artificial Intelligence）の活用への期待**が高まっており、現在は**第三次人工知能ブーム**とされています。
  - 第一次人工知能ブームは1960年前後、第二次人工知能ブームは1980年代であったとされています。
- 人工知能は、表のように**汎用人工知能**と**特化型人工知能**に大別することができます。

汎用人工知能と特化型人工知能の分類

分類	説明	イメージ・事例
汎用人工知能	**様々な思考・検討を行うことができ、初めて直面する状況に対応できる人工知能**	将棋、炊事、掃除、洗濯といった様々な分野および初めての状況に対する思考・検討ができる。
特化型人工知能	**特定の内容に関する思考・検討にだけに優れている人工知能**	• 将棋に関する思考・検討のみできる人工知能 • 掃除に関する思考・検討のみできる人工知能

- プロの棋士に勝てるほどに将棋に強い人工知能があっても、将棋以外に対応できない人工知能は、特化型人工知能に該当します。
- 「掃除のみ」「空調のみ」に関する思考・検討ができる特化型人工知能は、市販されている一部の家電製品に格納されています。

- 人工知能は、**強いAI**と**弱いAI**という分類もあり、**強いAIは意識や自我を持つAI**とされています。
  - 強いAIと弱いAIの分類は観念的であり、強いAIは概ね汎用人工知能に対応し、弱いAIは概ね特化型人工知能に対応すると示すケースもあります。
- 人工知能学会では、人工知能研究には**「人間の知能そのものをもつ機械を作ろうとする立場」**、**「人間が知能を使ってすることを機械にさせようとする立場」**の2種類があると示し、**実際の研究のほとんどは後者**と記しています。

【出所】人工知能って何？［(社) 人工知能学会］
https://www.ai-gakkai.or.jp/whatsai/AIwhats.html

人工知能に関する定義とイメージ

◆人工知能（AI）には、確立した学術的な定義や合意がありません。

- 専門家や研究者の間でも**「人工知能」に関する確立した学術的な定義、合意はありません。**
  - 平成28年版 情報通信白書（P234）では「国内の主な研究者による人工知能（AI）の定義」として13人の研究者による定義・説明を紹介していますが、その内容は様々です。

【出所】平成28年版 情報通信白書［総務省］
http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h28/pdf/n4200000.pdf

- 様々な人工知能の定義・説明の中には、**「知的」「知能を持つ」という言葉が含まれる**ケースがあります。
  ただし、「知的」「知能を持つ」という感覚は、**個々人の感じ方、考え方に依存する部分**もあります。
  - スマートフォンの音声アシスタントに「おはよう」と話して「おはようございます」と返事が返ってくることを「知能を持つ」と感じるかは、人それぞれです。

- 人工知能（AI）のイメージを尋ねた調査結果においても、日本とアメリカでは回答傾向が異なってなっており、**日本では会話を重視**しているのに対して、**アメリカでは認知・判断を重視**しています。

日米における人工知能（AI）のイメージに関する最大の回答割合の選択肢

調査への回答者［回答者数］	**複数回答における最大の回答割合となった選択肢**［回答割合］
日本の就労者［1,106人］	**コンピューターが人間のように見たり、聞いたり、話したりする技術**［35.6%］
アメリカの就労者［1,105人］	**人間の脳の認知・判断などの機能を、人間の脳の仕組みとは異なる仕組みで実現する技術**［42.3%］

【出所】ICTの進化が雇用と働き方に及ぼす影響に関する調査研究［総務省（調査委託先：株式会社 野村総合研究所）］
http://www.soumu.go.jp/johotsusintokei/linkdata/h28_03_houkoku.pdf

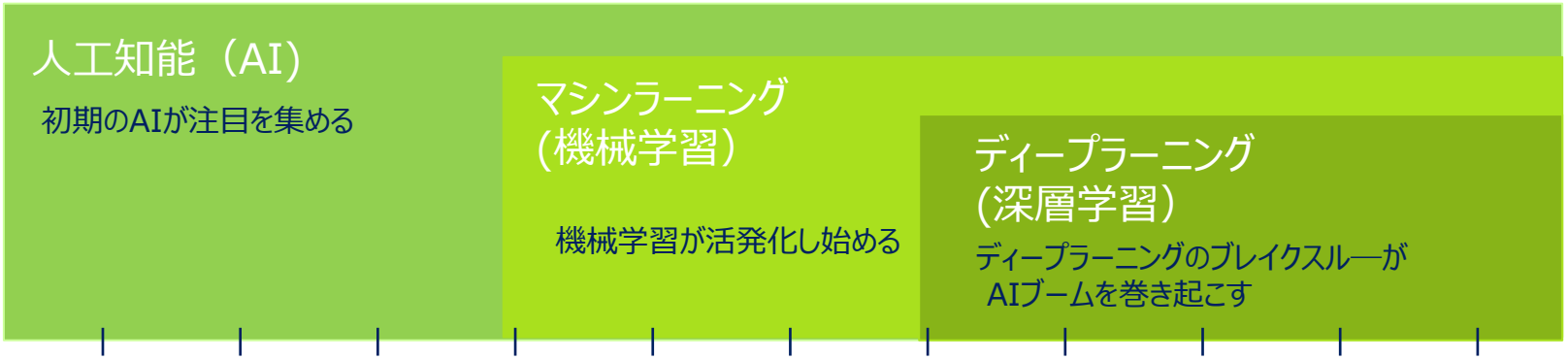
□ この講座では、人工知能の厳密な定義を行わず、それぞれの人の考え方に依存する部分を含めて「人間が知的と感じる情報処理・技術」と見なします。

人工知能に含まれる分析技術

◆「人工知能」「機械学習」「ディープラーニング」には包含関係があります。

- **人工知能に関わる分析技術として「機械学習」**が挙げられ、**機械学習の一つの技術として「ディープラーニング（深層学習）」**が挙げられます。
- **機械学習**とは**「データから規則性や判断基準を学習し、それに基づき未知のものを予測、判断する技術」**と**人工知能に関わる分析技術**を指しています。
  - 機械学習の研究初期には「学習する」点に注目されてきましたが、現在では「学習に基づいて予測・判断する」点に注目されることが多くなっています。
  - 機械学習の定義にも曖昧な面もあり、「データ分析」と概ね同義に使われるケース、人間にとって分析プロセスや判断基準が不明なものを含めて目的志向のデータ活用を強調するケースもあります。
- 近年、注目されている**統計学を応用する機械学習は、統計的機械学習**とも呼ばれます。
- **ディープラーニング（深層学習）**は、より基礎的で広範な機械学習の手法である**ニューラルネットワーク**という分析手法を拡張し、高精度の分析や活用を可能にした手法です。
  - 本講座内でニューラルネットワーク・ディープラーニングの分析イメージ、分析の特徴、活用されている用途を紹介します。

人工知能、機械学習、ディープラーニングの包含関係と隆盛



【出所】グーグルに学ぶディープラーニング［日経ビッグデータ編/日経BP社］に基づき作成

機械学習の分類

◆機械学習の分類として「教師あり学習」「教師なし学習」「強化学習」の3種が挙げられます。

- 機械学習は、真実のデータや人間による判別から得られた**正解に相当する「教師データ」**の与えられ方によって分類することができます。
  - 写真の画像から性別を分類する機械学習では、実際の性別や人間が行った判断が教師データとなります。

- 教師データの状況によって、機械学習は大きく、**教師あり学習**、**教師なし学習**、**強化学習**に分類されます。
  - 「教師あり学習」と「教師なし学習」は統計学に基づいた「統計的機械学習」が一般的な一方で、「強化学習」は、概ね統計学とは無関係です。
  - 教師データがある標本とない標本が組み合わさったデータを利用する「半教師あり学習」という種類もあります。
- **「教師あり学習」は、回帰や分類に利用**されるケースが多く、**「教師なし学習」はグループ分けや情報の要約に利用される**ケースが多くなっています。
- 強化学習では、**報酬（評価）**が与えられ、将棋のようなゲーム用の人工知能に応用されています。
  - ロボットの歩行距離に関する強化学習では、転倒せずに「歩けた距離」が報酬（評価）として与えられ、試行錯誤で歩行距離を伸ばします。

利用可能なデータに基づく機械学習の分類

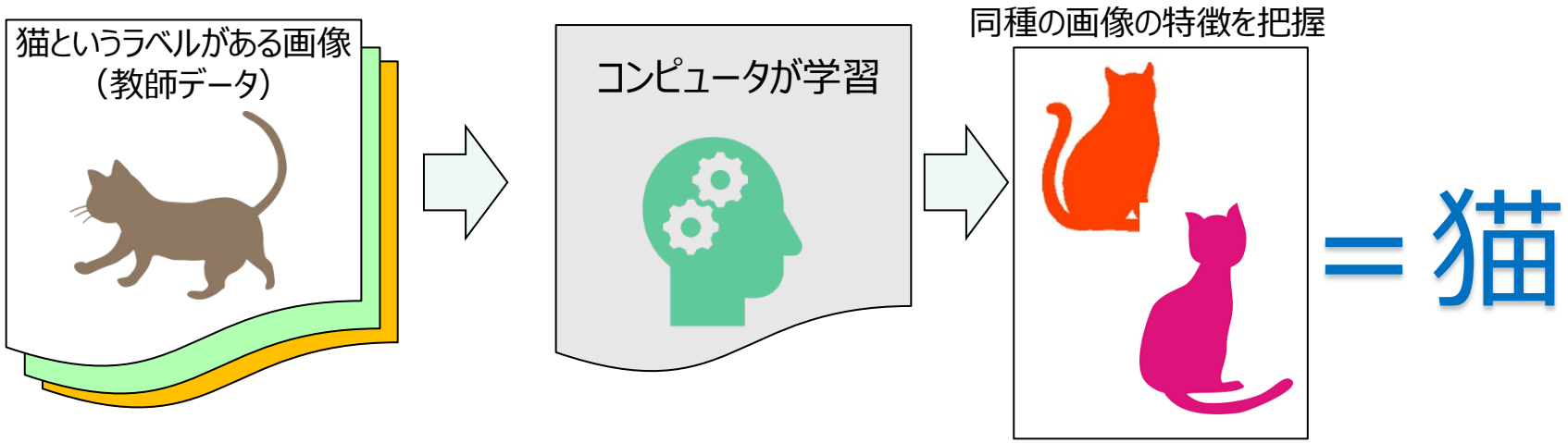
	入力に関するデータ［質問］	出力に関するデータ（教師データ）［正しい答え］		主な活用事例
教師あり学習	与えられる	○	与えられる	出力に関する回帰、分類
教師なし学習	与えられる	×	与えられない	入力に関するグループ分け、情報の要約
強化学習	与えられる（試行する）	△（間接的）	正しい答え自体は与えられないが、報酬（評価）が与えられる	将棋、囲碁、ロボットの歩行学習

教師あり学習

◆教師あり学習は正解に相当する教師データが与えられ、主に回帰や分類に利用されます。

- **「教師あり学習」**は、**結果や正解にあたる「教師データ」が与えられるタイプの機械学習**です。
  - 写真から年齢・性別を判定するモデルでは、戸籍等に基づく年齢・性別を「教師データ」とすることもあれば、人間が見て判別した年齢・性別を「教師データ」とするケースもあります。
- 教師あり学習では、**教師データを既知の情報**として学習に利用し、未知の情報に対応することができる**回帰モデル**や**分類モデル**を構築します。
  - 回帰は講座3-4で示した回帰分析のように出力側の被説明変数として連続値を扱う一方で、分類の出力側は「優/良/可/不可」といった評価や移動手段の「徒歩/バス/タクシー」のような離散値です。
- 例えば、「猫」というラベル（教師データ）が付けられた大量の写真をコンピュータが学習することで、ラベルのない写真が与えられても、「猫」を検出できるようになります。
  - コンピュータによる写真のラベル付けは、人間があらかじめ画像に割り当てたラベルを教師データとする「教師あり学習」に該当します。

教師あり学習のイメージ

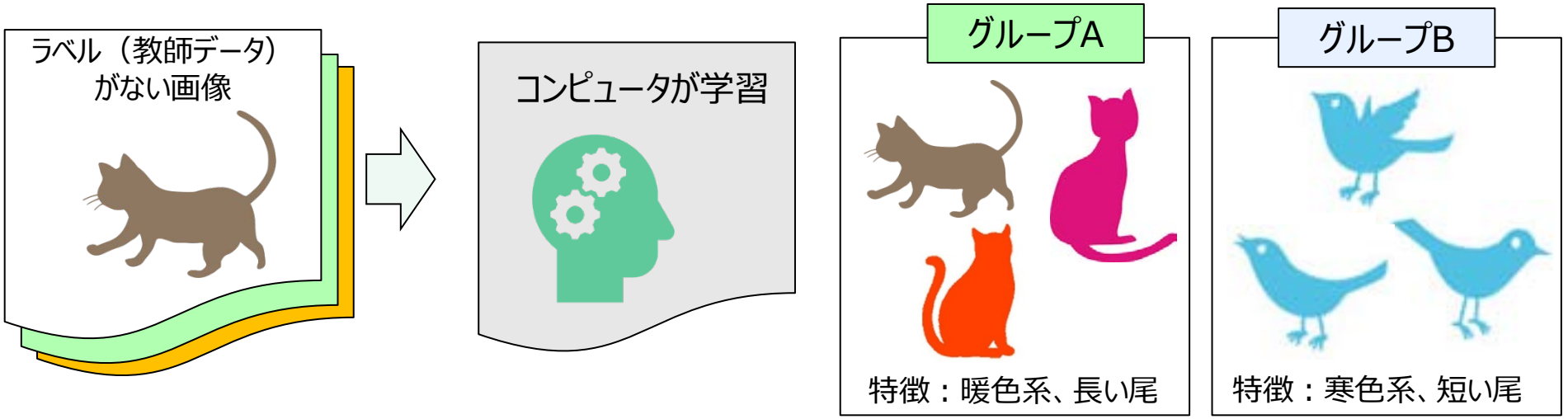


教師なし学習

◆教師なし学習には教師データがなく、データのグループ分けや情報の要約に利用されます。

- **「教師なし学習」**は、**正解に相当する「教師データ」が与えられないタイプ**の機械学習です。
  - 「教師あり学習」では、現実のアウトプットに関するデータや人間が判別して与えた正解に相当する「教師データ」が与えられていました。
  - アウトプットに関するデータや正解に相当するデータ・ラベルは与えられず、インプットに関するデータのみであっても、教師なし学習は対応可能です。
- 教師なし学習では、データの**グループ分け**や**情報の要約**などに活用されます。
  - データのグループ分けは「クラスタリング」と呼ばれる分析手法が代表的であり、情報の要約は「次元圧縮」という分析手法が代表的です。
- 教師データに相当するラベルがない場合であっても、大量の画像をコンピュータに学習させれば、画像の特徴（例：大きさ、色、形状）からグループ分けや情報の要約が可能です。
  - 教師なし学習では、「猫」や「鳥」というラベルは与えられていませんが、形や色などが近い属性でグループ分けできます。
  - コンピュータがグループの名前をつけることはできず、「グループＡ」「グループＢ」といったラベルがないグループになります。

教師なし学習のイメージ

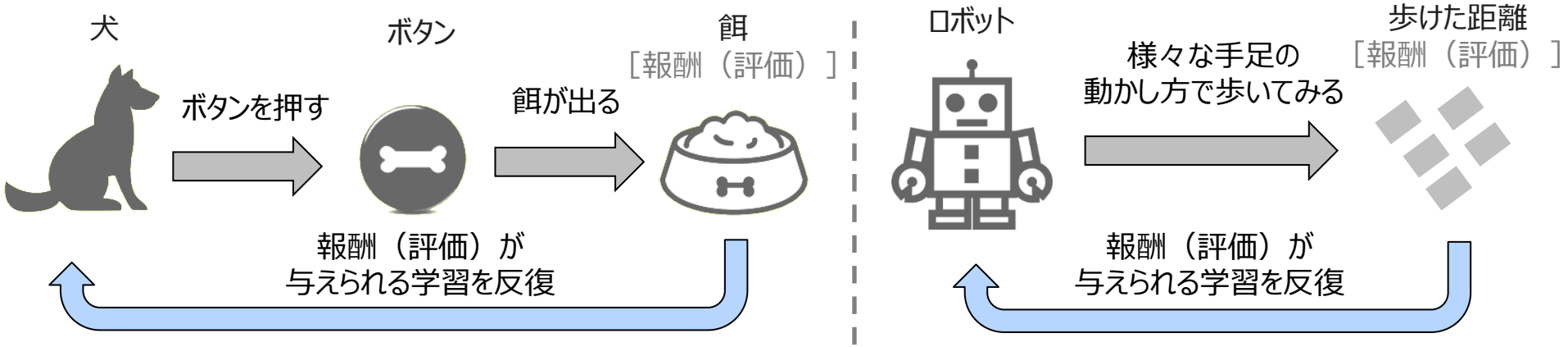


強化学習

◆強化学習では、試行錯誤を通じて、報酬（評価）が得られる行動や選択を学習します。

- 強化学習の例として、犬がいる部屋に「ボタンを押すと餌が出る機械」を設置した場合を考えます。
  - 犬が偶然にボタンを押すと餌が出ますが、1度だけの経験ではボタンと餌が出たことを結びつけられず、犬には餌が出た理由が分かりません。
- 「ボタンを押す」という試行を犬が繰り返すと、犬は「ボタンを押す⇒餌が出る」ということを学習します。
  - 犬にとっては「ボタンを押す」という行動が入力であり、「得られる餌」が報酬（評価）に当たります。
- ロボットの歩行に関しても、「歩けた距離」を**報酬（評価）**として設定し、手足の動かし方を試行錯誤して歩行距離を伸ばすことが、**強化学習**に相当します。

強化学習のイメージ



- 将棋に関する強化学習は、敵軍の王将をとることに最大の報酬（評価）として設定し、コンピュータに評価を高める指し方を反復して学習させます。
  - 将棋では、詰め将棋を除いて正解の一手という教師データはありませんが、勝利した場合に最大の報酬（評価）を与え、勝利に近い局面ほど高い評価を与えることが可能です。

□ 強化学習は、一般に統計的機械学習の範疇外にあるため、以降では「教師あり学習」「教師なし学習」に絞って、統計的機械学習における分析手法を示します。

機械学習の代表的な分析手法

◆「教師あり学習」「教師なし学習」には、それぞれ代表的な分析手法があります。

- 代表的な**教師あり学習**の分析手法として、**回帰分析**や**決定木**などが挙げられます。
  - 講座3-4でExcelの分析ツールでの実行方法を示した回帰分析は「教師あり学習」の一つの手法として位置づけられます。

代表的な「教師あり学習」の分析手法

分析手法名	分析手法・用途の概要
回帰分析	被説明変数と説明変数の関係を定量的に分析し、分析結果に基づく予測
決定木	分類のための基準（境界線）を学習し、未知の状況でデータを分類

- 代表的な**教師なし学習**の分析手法として、**k平均法**、**アソシエーション分析**が挙げられます。

代表的な「教師なし学習」の分析手法

分析手法名	分析手法・用途の概要
k平均法	特徴・傾向が似ている標本をいくつかのグループに分類
アソシエーション分析	同時購入の頻度等を算出し、消費者の選択・購入履歴から推薦すべき商品を導出
ソーシャルネットワーク分析	氏名が同時掲載される頻度やSNS上での友人関係から人のつながりを分析

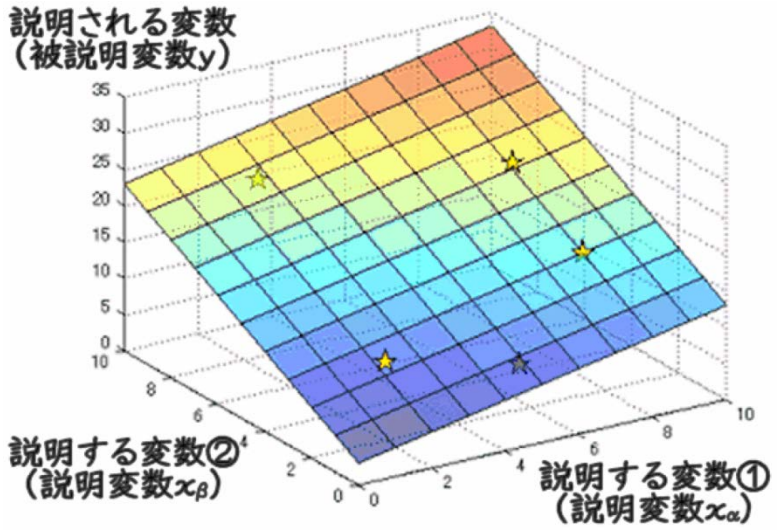
- 近年、注目されている「ディープラーニング（深層学習）」は、教師なし学習、強化学習への応用もあり得ますが、教師あり学習としての活用が一般的です。

回帰分析（最小二乗法）

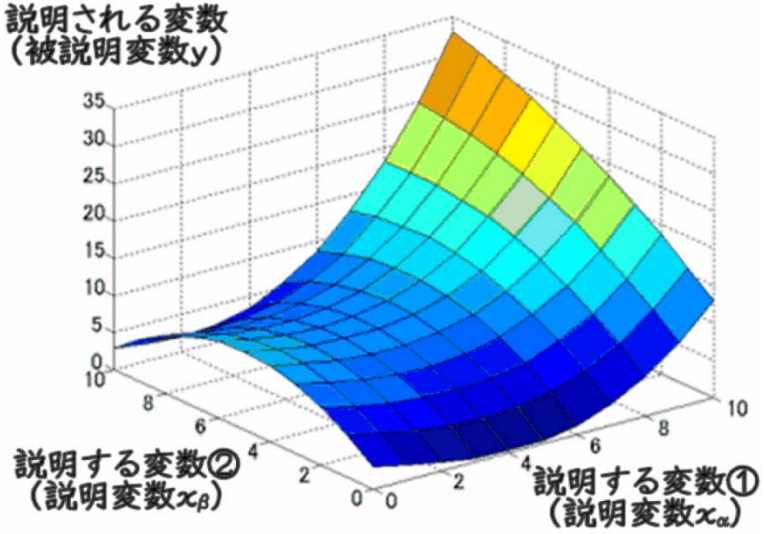
◆回帰分析（最小二乗法）は、教師あり学習の一手法でもあります。

- 講座3-4において、Excelの分析ツールによる手順を示した**回帰分析（最小二乗法）**も**機械学習の一種**です。
  - 回帰分析はデータの規則性に基づいて予測を行うことができ、機械学習の概念に含まれます。
- 回帰分析は、アウトプットに関するデータである**被説明変数を教師データとして利用**し、**教師あり学習**に分類されます。
  - 被説明変数と説明変数の間の規則性を回帰式として導出した後は、説明変数に新たな値を与えても被説明変数の予測ができます。
- スーパーマーケットの売上額（連続値）は、「チラシの配布数」「曜日」「天候（気温・湿度・天気）」によって説明されるという回帰式を作ることができます。
  - 過去のデータから、これらの変数間の関係を回帰式で導出すれば、「チラシの配布数」「曜日」「天気（予報）」が与えれば、「売上高」を予測することができます。

回帰分析の分析イメージ（線形）



回帰分析の分析イメージ（非線形）

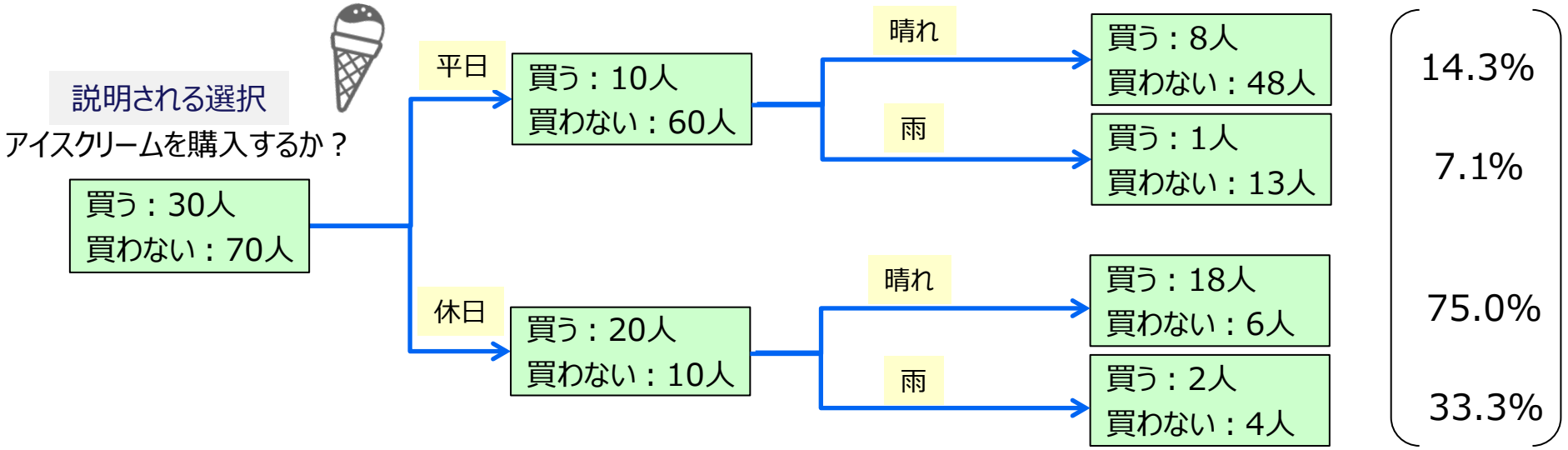


決定木

◆購入の有無などの結果（教師データ）をもとにデータを分類し、様々な要因が結果に与える影響を把握する教師あり学習の一手法です。

- **決定木**（けっていぎ）は、木の枝のような段階を経て分かれる形（**樹形図**：じゅけいず）で判別基準を設定し、データを**分類**する手法です。
  - 決定木は、現実における選択や分類を教師データとして与える「教師あり学習」の一種です。
- 決定木における教師データは「商品を買う/買わない」「○円分購入する」といった**選択**であることが一般的です。
  - 「商品を買う/買わない」といった区分を分析する場合は分類木（ぶんるいぎ）、「○円分購入する」という連続的に変化しうる値を分析する場合は、回帰木（かいきぎ）と言います。
- アイスクリームの購入が、休日や天候から影響を受ける場合は、その影響の度合いを決定木で分析できます。
  - 決定木によって、どのような状況や要素が、利用者の行動を変化させるかを判別することができます。

決定木の樹形図（アイスクリームの購入選択）



k平均法（クラスタリング）

◆各標本を似たもの同士のグループ（クラスター）へ分類する教師なし学習の一つです。

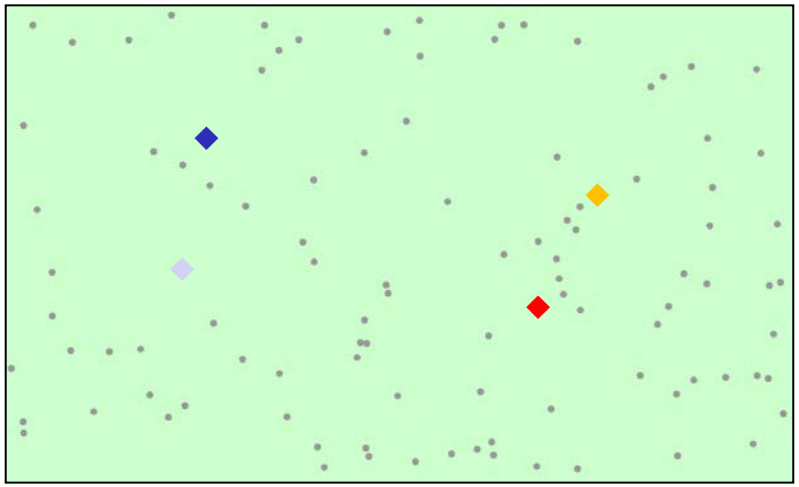
- **k平均法**は標本を**グループ分け**する手法で、教師データのない**教師なし学習**です。
  - データ分析におけるグループは、クラスター（群）という専門用語が使われるため、グループ分けの手法はクラスタリングと呼ばれます。

K平均法の手順

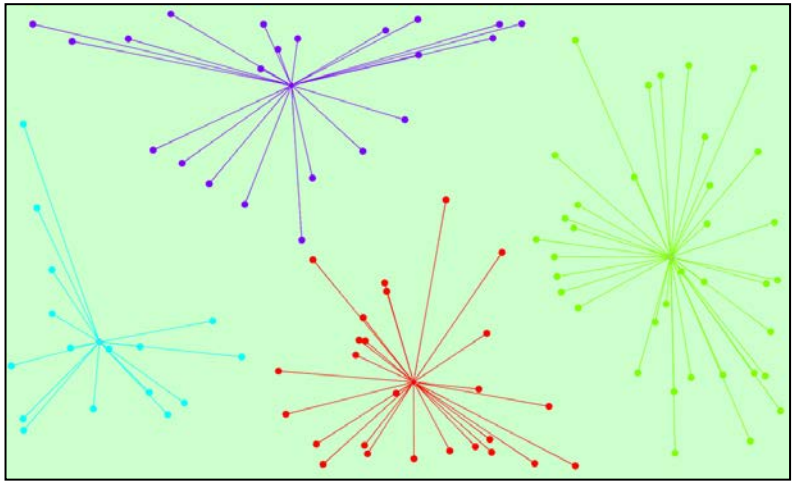
(1) 分類するグループ数をk個に定め、グループの中心となる点として、k個の◆の位置をランダムに与える。
(2) 各標本●は、一番距離が近い◆に属すると考えて、 各◆に属する標本●にグループ分けする
(3) グループ毎の●の座標の平均値（重心）をとって、その重心の位置に◆を移動させる。
(4) 移動させた後の◆に一番距離が近い●を取り直すことで、再び 各◆に属する標本●にグループ分けする
［◆が動かなくなるまで（3）と（4）を繰り返す］

動きが分かるデモサイト
https://wish.to/km

K平均法の初期値



K平均法によるグループ分けの結果



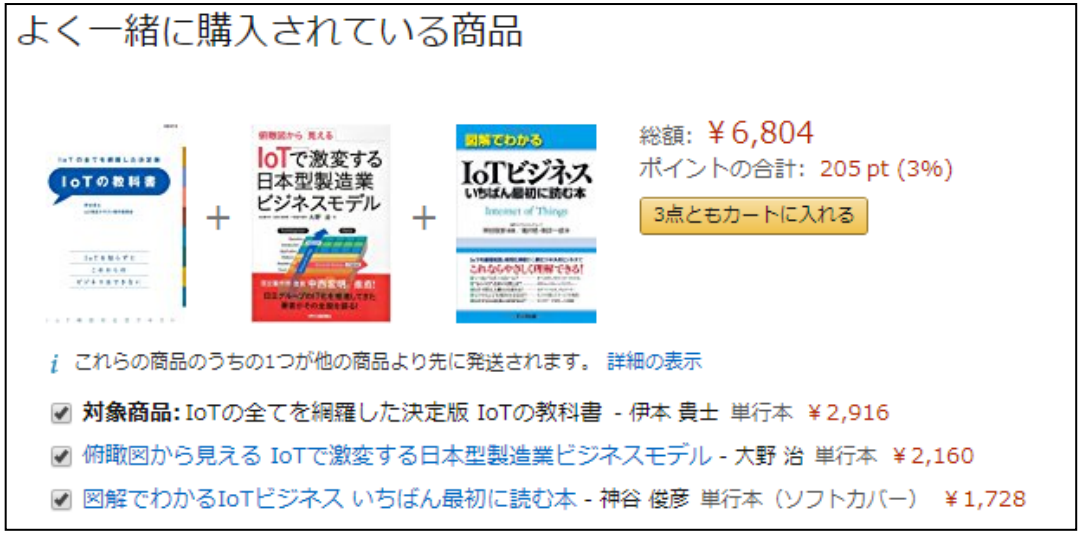
- グループ分け（クラスタリング）を用いることで、複数の観点から顧客や商品を分類することができます。
  - 「安さを重視する消費者」「高級感を重視する消費者」とグループを分けることができれば、グループに応じた販促活動に活用できます。

アソシエーション分析

◆同時に購入される商品セットやその確率を算出する教師なし学習の一つで、ネットショッピングサイトの推薦商品の提示にも利用されています。

- **アソシエーション分析**は、「商品Aを買っている人の○%が商品Bも買っている」といった**同時購入の確率**などを導出する手法です。
  - アソシエーション分析の主な評価指標として「支持度（同時確率）」「確信度（条件付き確率）」「リフト値（改善率）」の3種が挙げられます。
  - アソシエーション分析は、講座3-4で示した相関係数とは異なり、3種以上の商品の組み合わせに関しても分析できます。
- アソシエーション分析で把握できた同時購入等の情報は、ネットショッピングにおける**推薦（レコメンデーション）**や実店舗における商品の陳列やセット割引きの検討にも利用できます。
  - Amazon.co.jpをはじめとするネットショッピングのサイトでは、「この商品を買った人はこんな商品も買っています」等のメッセージで、関連商品の購入を勧める手法でも活用されています。

Amazonの同時購入商品の推薦



【出所】Amazon.co.jp https://www.amazon.co.jp/

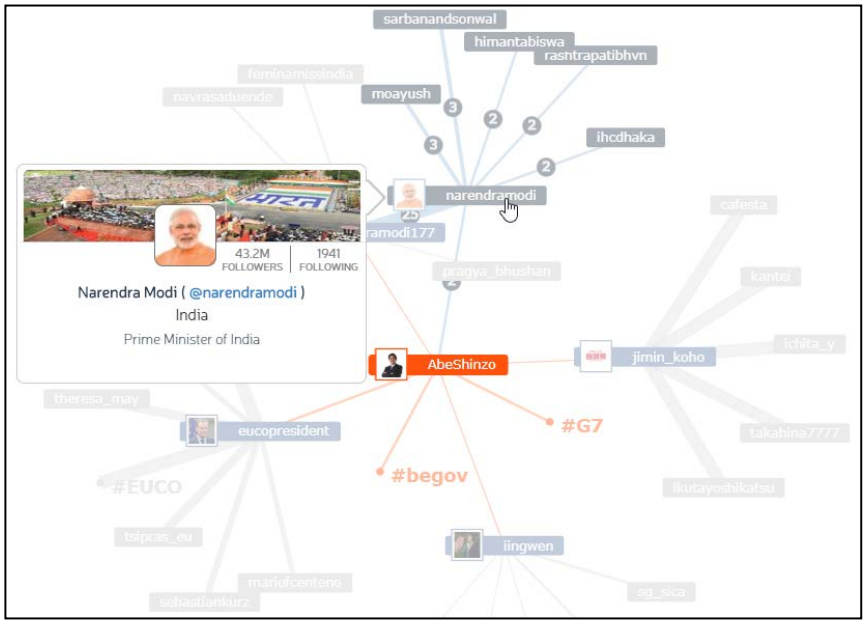
ソーシャルネットワーク分析

◆ソーシャルネットワーク分析は、人間のつながりを可視化する教師なし学習の一つです。

- 公表資料内で氏名が同時掲載される頻度やSNS上で友人としてのつながりのデータに基づき、人のつながりの分析手法として**ソーシャルネットワーク分析**が挙げられます。

- Mentionmapp Analytics ではTwitter（ツイッター）の情報に基づく人間関係を表せます。
  - 安倍首相をはじめとして、現在Twitterのアカウントを公に利用している人に関する分析ができます。

安倍首相のTwitterに関する
Mentionmapp Analyticsの可視化



【出所】Mentionmapp Analytics
https://mentionmapp.com/

- Twitterにおける関係の強さ、よく使う言葉を可視化できるwhotwi（フーツイ）というサービスもあります。
  - 安倍首相のTwitterアカウントは、インドのモディ首相のアカウントとの関係が強いことが示されています。

安倍首相のTwitterに関する
whotwiの可視化

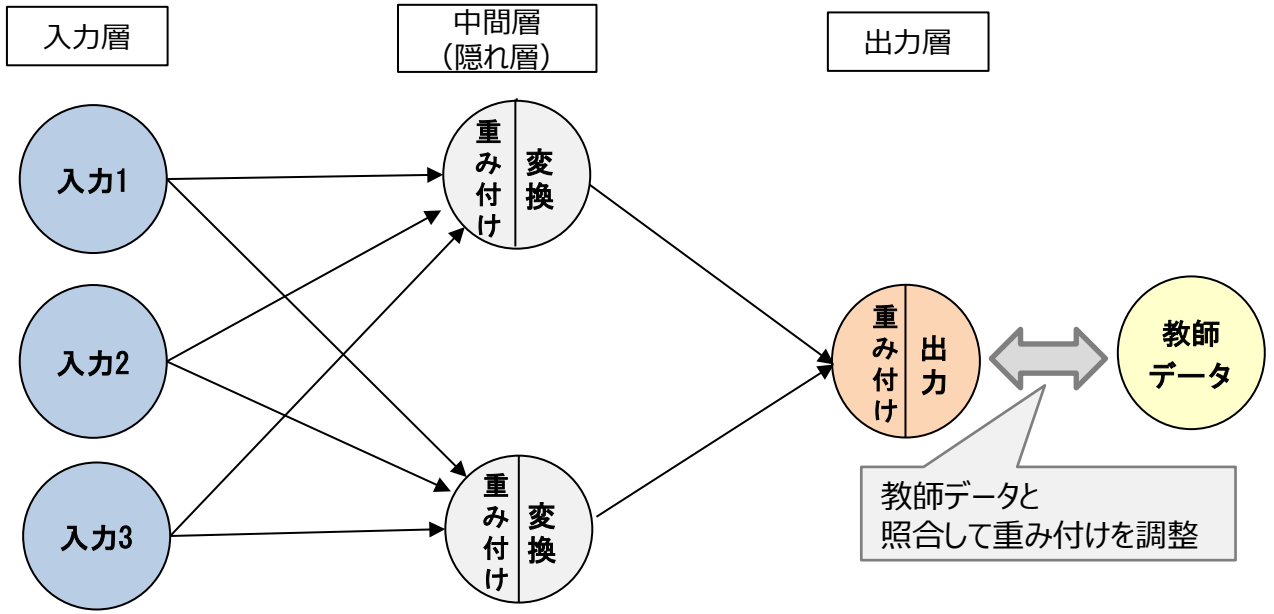
【出所】whotwi
https://ja.whotwi.com/abeshinzo

ニューラルネットワーク

◆ニューラルネットワークは、文字や音声の認識といったパターン認識へ応用されています。

- **ディープラーニング**の基本となる**ニューラルネットワーク**は、**脳の神経回路の仕組み**を模した分析モデルです。
  - ニューラル（neural）は「神経の」という意味があります。
- ニューラルネットワークは**入力層**、**中間層（隠れ層）**、**出力層**の3層から成り立ちます。
  - 中間層（隠れ層）では、一つ前の層から受け取ったデータに対し「重み付け」と「変換」を施して次の層へ渡します。
- ニューラルネットワークの出力は教師データ等と照合され、より一致度が高くなるように**重みのつけ方を調整**します。
  - ニューラルネットワークは、教師なし学習や強化学習にも応用されますが、教師あり学習としての利用頻度が高くなっています。

ニューラルネットワーク（中間層が1層のケース）のイメージ



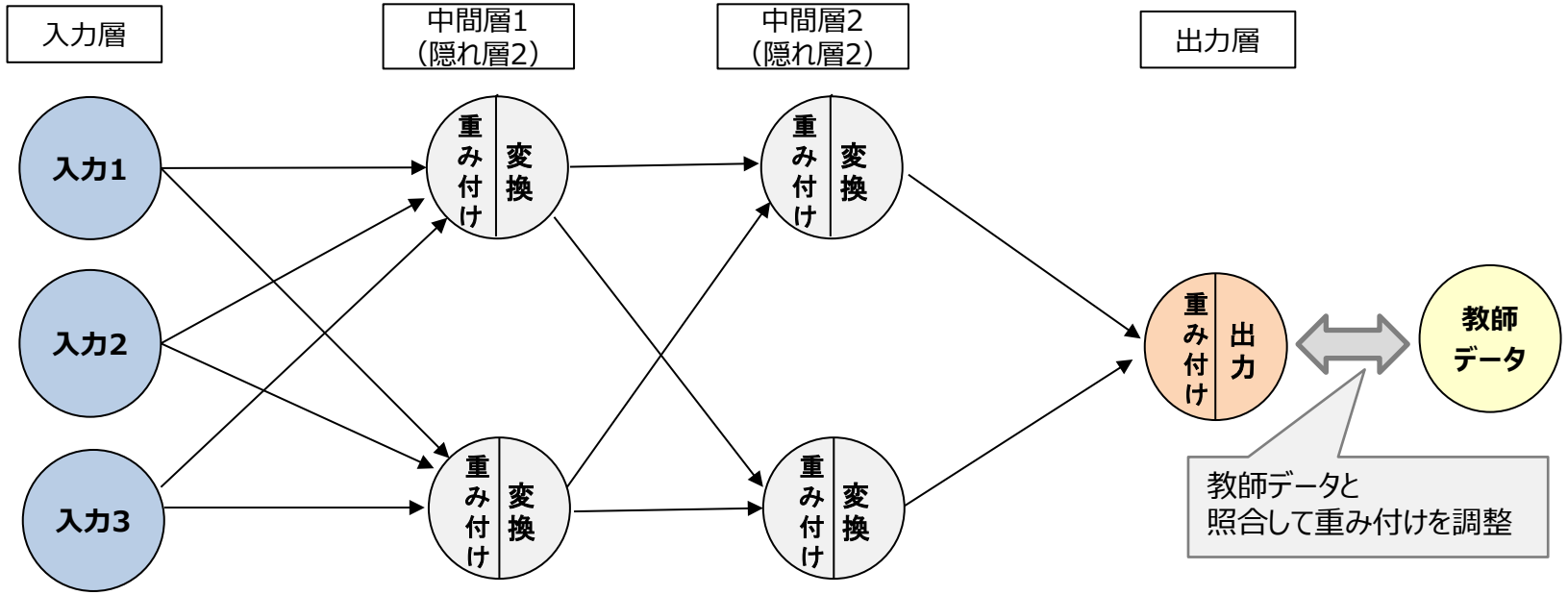
- ニューラルネットワークは回帰、分類、画像認識、音声認識、翻訳といった様々な分野で応用されています。

ディープラーニング（深層学習）

◆ディープラーニング（深層学習）は、より高精度の認識などに活用されています。

- **中間層（隠れ層）を2層以上に多層化したニューラルネットワーク**を**ディープラーニング**といいます。
  - ディープラーニングの日本語訳は「深層学習」ですが、この「層」と言う言葉は「中間層が複数あること」を示しています。
- 中間層が複数あることで中間層が1層の場合に比べて、**より教師データに合致する複雑な出力**をすることができます。
  - ディープラーニングの原理はニューラルネットワークと同じですが、中間層が多層化することでその精度が向上しました。
  - ニューラルネットワークは中間層が複数あるケースを含み、ディープラーニングはニューラルネットワークの一種です。

ディープラーニング（中間層が2層）のイメージ



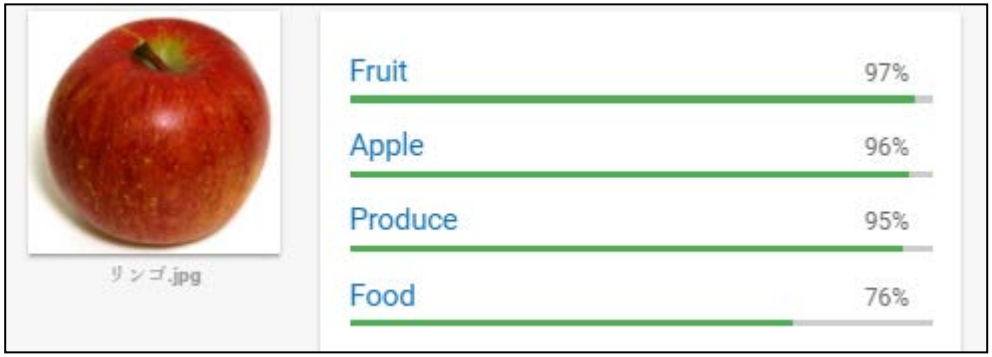
□ 中間層（隠れ層）を多層化したディープラーニングは、教師データとより高精度に対応づけることが可能になりました。

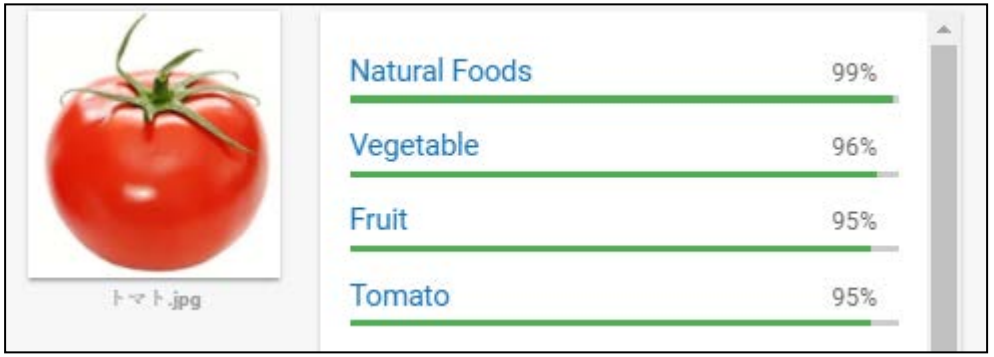
ニューラルネットワーク・ディープラーニングの特徴

◆ニューラルネットワーク・ディープラーニングは、従来型のプログラムと異なる特徴があります。

- ディープラーニングによって、**特徴量と呼ばれるデータの中で注目すべきポイント**をコンピュータ自ら検出できるようになりつつあります。
  - 画像データをプログラミングによって分類する場合、従来は「リンゴは赤い」「リンゴは丸い」といった特徴量を人間が入力、指定する必要がありました。
- 人間がプログラミングすることなく、大量の教師データを与えることで、**コンピュータ自ら学習**することができます。
  - 従来型の「If-Then-Else型」のプログラムでは「丸くて赤い物であれば、リンゴ」と出力することはできますが、同じ丸くて赤い物である「リンゴとトマトの識別」の条件設定やプログラミングは困難です。
  - ニューラルネット・ディープラーニングは従来型のプログラムとは根本的に異なり、論理や言語で説明困難な特徴をコンピュータが自動的に習得します。
- コンピューターの判別基準は、サービス提供者にとってもブラックボックスであり、説明することができません。
  - リンゴとトマトを高い精度で識別・ラベル付与ができますが、なぜそのように識別したのかは説明できません。
  - 識別・ラベル付与に関する信頼度合いを出力できますが、識別の基準（例：色が3割、形が7割）といった分解はできません。
- ニューラルネット・ディープラーニングは、本番同様の学習・環境設定を行わずには精度や実用性の想定は困難です。

GoogleのCLOUD VISION APIにおける画像認識のラベル付け（リンゴとトマト）





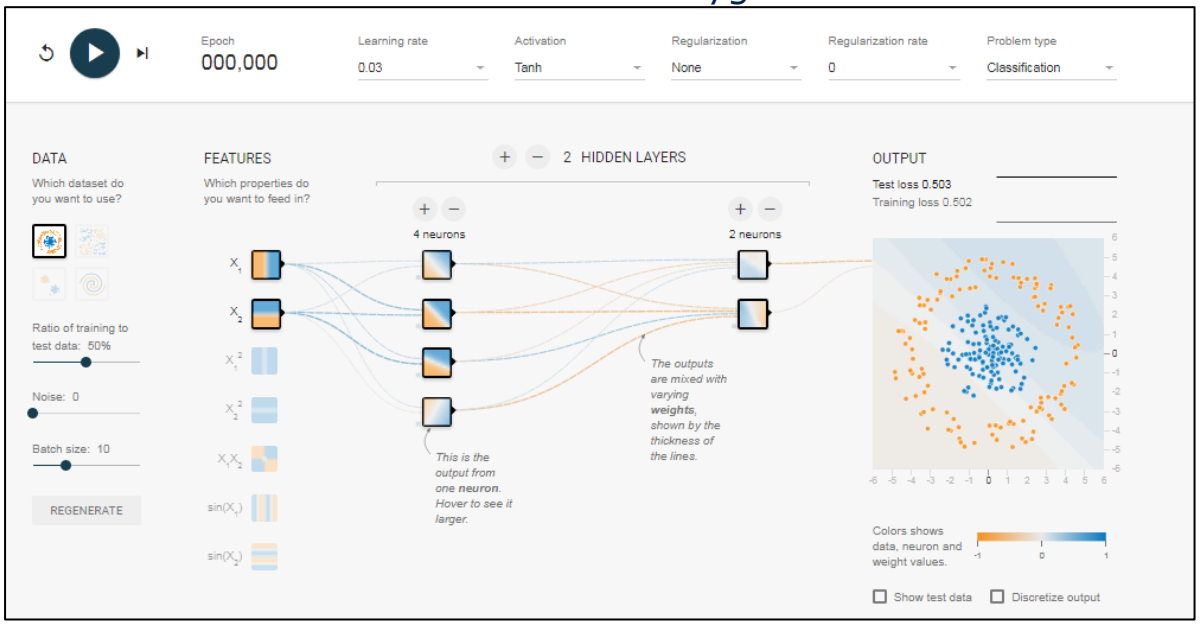
【出所】CLOUD VISION API［Google Cloud Platform］https://cloud.google.com/vision/?hl=ja

A Neural Network Playground

◆「ニューラルネットワーク」「ディープラーニング」のイメージを確認できるウェブサイトもあります。

- 「A Neural Network Playground」は、画面左上の再生ボタンを押すことで、画面右側の青とオレンジの点を塗り分けを行う「ニューラルネットワーク」「ディープラーニング」の確認サイトです。
  - 「A Neural Network Playground」は、Daniel Smilkov氏とShan Carter氏によって、主に教育目的で制作されました。
  - 右側の画像の上にある「Test loss」や「Training loss」は、それぞれテスト用データと学習用データで正しく塗り分けることができなかった割合を指します。
  - 機械学習においては、入手したデータをモデル構築用の学習用データと構築したモデルの精度を検証するテスト用データに分けることがあります。
- 再生ボタンを押して10秒程度待つと、学習用データに合わせて重みを調整することで、loss（ロス）が小さくなります。
  - 画面上に[Epoch]と表示されているのが、教師データに合わせて重みを調節し直した回数に該当し、再生時間に応じて増加します。
- 画面左側の画像で模様を複雑にしたり、Noise（ノイズ）を大きくすることで、塗り分けをより難しくできます。

「A Neural Network Playground」の画面



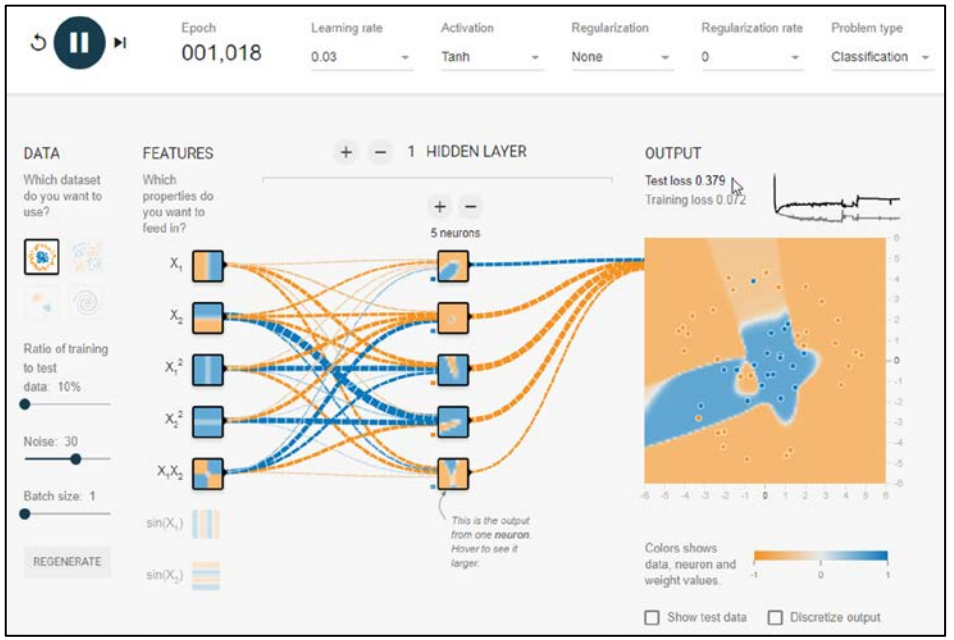
【出所】A Neural Network Playground http://playground.tensorflow.org/

隠れ層が1層の場合（ニューラルネットワークのイメージ）

◆隠れ層が一層のケースが、ディープラーニングではないニューラルネットワークに該当します。

- HIDDEN LAYER（隠れ層）が1層となっていると、ディープラーニングではないニューラルネットワークのイメージです。
  - HIDDEN LAYER（隠れ層）の左側の「＋-」をクリックすることで、中間層を増減させることができます。
- 下の画像の例では、初期設定からノイズを増加させ30とし、各層におけるユニット数は5個としています。
  - 各層の上側にある「＋-」をクリックすることで層内の神経細胞にあたるユニット数を増減させることができます。

隠れ層が1層のニューラルネットワークのイメージ



【出所】A Neural Network Playground http://playground.tensorflow.org/

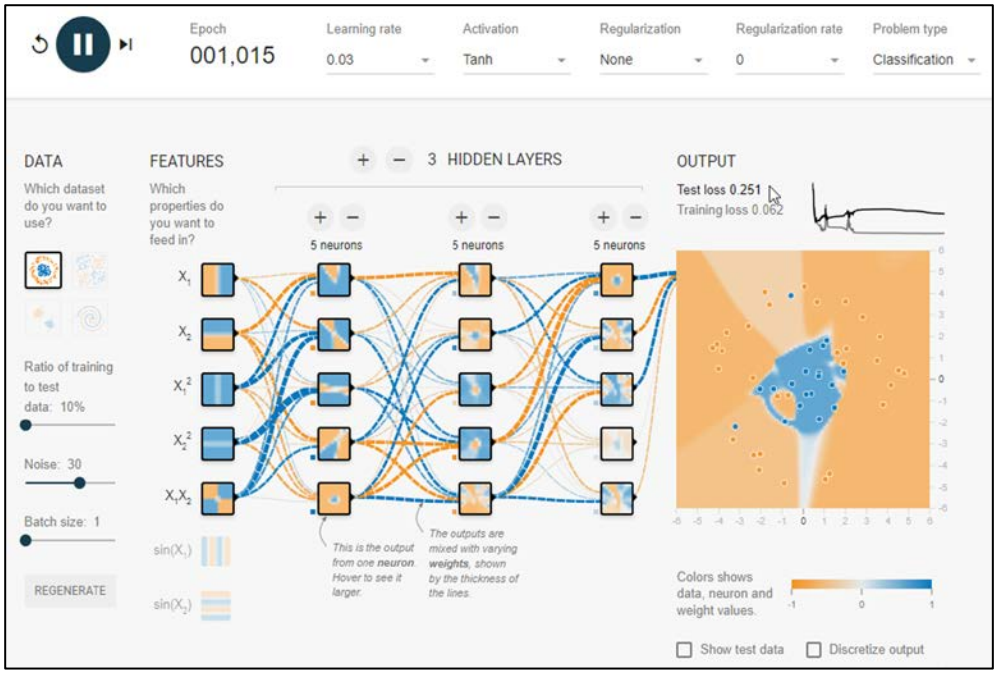
- 約1000回重みの調整をした段階におけるTest lossが0.379、Training lossが0.072と算出され、画面右の画像により塗り分けられない部分の割合を確認できます。
  - A Neural Network Playgroundでは、画面左下の「REGENERATE」をクリックすると点が再生成されるため、実行ごとに微妙に結果は異なります。

隠れ層が2層以上の場合（ディープラーニングのイメージ）

◆隠れ層を増やし、ディープラーニングに対応する形にすると、lossが低下する傾向があります。

- HIDDEN LAYER（隠れ層）が2層以上となっているのが、ディープラーニングのイメージです。
- 下の画像の例では、隠れ層を3層として、各層のユニット数はニューラルネットワークと同じ5としています。

隠れ層が3層のディープラーニングのイメージ



【出所】A Neural NetworkPlayground http://playground.tensorflow.org/

- 約1000回重みの調整をした段階におけるTest lossが0.251、Training lossが0.062と算出され、塗り分けられない部分の割合は隠れ層が1層のときと比較して減っていることが分かります。

□ ディープラーニングを利用することで、より高精度の分析や判別が可能となってきています。

人工知能サービスに関するクラウド・APIの利用

◆クラウド・APIを利用することで、安価に高精度の人工知能サービスを利用できます。

- 「教師あり学習」の人工知能サービスの精度を高めるためには、より多くの教師データが必要です。
  - 教師データの形式が不統一であったり、教師データに誤りがあったりすると、高精度の人工知能サービスの構築ができません。高精度の人工知能サービスの構築には、まず人間が講座3-1に示した「データ形式を標準化」「データクレンジング」を行うことで、教師データを整備することが必要です。
- 世界規模のクラウドプラットフォームでは、相対的に高精度の人工知能サービスを提供しています。
  - 講座2-2に示したAmazonが運営するAWS（Amazon Web Service）、Microsoftが運営するAzure、Googleが運営するGoogle Cloud Platformの3大クラウドプラットフォームに加えて、IBMが運営するIBM CloudのWatsonでは、世界規模で人工知能サービスを提供しています。

世界規模でのクラウドプラットフォーム・人工知能サービスを提供しているブランド

Amazon Web Servicesは、米国その他の諸国における、Amazon.com, Inc.またはその関連会社の商標です。

Microsoft Azureは、米国Microsoft Corporation およびその関連会社の商標です。

© 2017 Google LLC All rights reserved. Google および Google ロゴは Google LLC の登録商標です。

IBM Watson is a trademark of International Business Machines Corporation, registered in many jurisdictions worldwide.

- 世界規模のクラウドプラットフォームは、検索サービスを提供しているなど、人工知能サービスの精度向上に必要な教師データを収集しやすい環境を持っています。
- クラウドが提供している人工知能サービスを一時的に借りる方法として、APIの利用が挙げられます。
  - 講座1-5において示したように、インターネットを経由して人工知能に関するサービスを利用できるAPIもあります。
  - ディープラーニングは、並列計算による高速化の余地が大きく、講座2-5に示したGPUクラウドコンピューティングも活用されています。

ニューラルネットワーク・ディープラーニングの応用例

◆ニューラルネットワーク・ディープラーニングのビジネス面への応用には代表例があります。

- ニューラルネットワーク・ディープラーニングを用いた人工知能サービスの代表例として、**画像認識**、**音声認識**、**音声合成**、**テキスト処理**、**翻訳**が挙げられます。

画像認識	画像データから、連想されるキーワードを出力したり、顔の画像を検出したりできる。 応用例： • 顔画像から性別や年齢を推定 • 手書き文字画像のテキスト変換	**テキスト処理**	テキストデータから、内容を要約したり、内容に基づく文章の採点・分類をしたりできる。 応用例： • 議事録から議事要旨の作成 • 口語体から文語体への変換
音声認識	音声データから、人間の発声を言葉として認識し、テキストとして文字起こしできる。 応用例： • 声の抑揚に基づく強調点の抽出 • 声から健康やストレスの検出	**翻訳**	テキストデータから、経験豊かな翻訳者が訳したような正確かつ自然な翻訳文を作成できる。（多言語翻訳にも対応） 応用例： • 特定の翻訳者の訳し方を反映した翻訳
音声合成	テキストデータから、自然に聞こえる人間の発声を合成し、文章として読み上げることができる。 応用例： • 歌を歌う（ボーカロイド） • 特定の人物の発声を真似る	**複数機能の組み合わせによる高度な活用も可能** ◆「画像認識＋音声認識」を組み合わせることで、複数の知覚に基づく高精度の認識 ◆「音声認識→テキスト処理→音声合成」を組み合わせることで、人間と対話できるコンピューターの構築	

ニューラルネットワーク・ディープラーニングのサービス例

◆ニューラルネットワーク・ディープラーニングは、画像認識や翻訳に活用されています。

- Microsoft Azureでは、ディープラーニングに基づく画像認識をウェブブラウザを使って体験できます。
  - AzureのComputer VISION APIのデモでは、利用者が指定したウェブ上の画像やアップロードした画像に対して連想されるキーワードを表示します。

Microsoft Azureのデモサイトにおける画像認識

特徴名:	値
説明	{ "tags": ["outdoor", "building", "city", "background", "large", "water", "tall", "clock", "front", "tower", "top", "sitting", "street", "boat", "parked", "standing", "blue", "river", "sign", "ocean", "flying", "airplane", "man"], "captions": [{ "text": "the tower of the city", "confidence": 0.9085732 }] }
タグ	[{ "name": "sky", "confidence": 0.99568224 }, { "name": "outdoor", "confidence": 0.990158 }, { "name": "building", "confidence": 0.96608603 }, { "name": "city", "confidence": 0.5801872 }, { "name": "tall", "confidence": 0.4686955 }, { "name": "tower", "confidence": 0.343788415 }]

東京スカイツリーの画像を指定すると、
sky（空）
outdoor（野外）
building（建物）
city（都市）
tall（高い）
tower（塔）
といった連想されるキーワードを出力します。

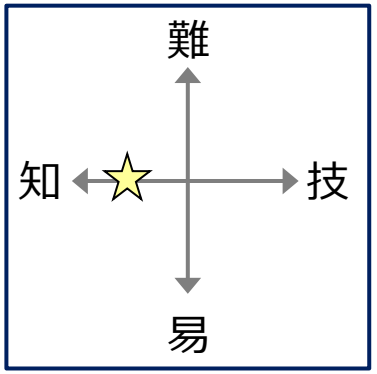
【出所】Computer VISION API［Microsoft（Azure）］
https://azure.microsoft.com/ja-jp/services/cognitive-services/computer-vision/

- Google翻訳は、2016年11月の「ニューラルネットに基づく機械翻訳 (Neural Machine Translation)」の導入により翻訳精度が向上しました。
  - Googleの翻訳チームは「過去10年の開発の歴史を振り返っても、それを大きく上回る飛躍的な前進」と記しています。

【出所】Google 翻訳が進化しました。［Google Japan Blog］
https://japan.googleblog.com/2016/11/google.html

□ 人工知能の分析技術のニューラルネットワーク・ディープラーニングは、様々なサービスにおける活用が期待されます。

総務省　ICTスキル総合習得教材



［コース4］オープンデータ・ビッグデータ利活用事例

4-1：オープンデータの利活用

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_4_1.pdf

	1	2	3	4	5
［コース1］データ収集					
［コース2］データ蓄積					
［コース3］データ分析					
［コース4］データ利活用	●				

本講座の学習内容［4-1：オープンデータの利活用］

【講座概要】

- オープンデータの事例、日本のオープンデータ基本指針におけるオープンデータの定義を説明します。
- 機械判読の適性の5段階など、国際的なオープンデータに関する取り組みを紹介します。
- オープンデータの関連概念であるクリエイティブ・コモンズとパブリックドメインについて説明します。
- オープンデータの相互運用性による複数のデータを組み合わせた活用例を示します。

【講座構成】

座学

- [1] オープンデータと日本における取り組み
- [2] 国際的なオープンデータに関する取り組み
- [3] オープンデータとクリエイティブ・コモンズ
- [4] 相互運用性とオープンデータの利活用

【学習のゴール】

- ✓ オープンデータ基本指針におけるオープンデータの定義を理解する。
- ✓ クリエイティブ・コモンズの各ライセンス、パブリックドメインについて理解する。
- ✓ オープンデータの相互運用性と実際の利活用例を紹介できるようになる。

様々な「オープンデータ」の活用

◆様々なレベルでの「オープンデータ」の活用によって、ウェブサイトやアプリが作られています。

- ウェブサイト「税金はどこへ行った？」では、各自治体が公表している**予算に関するオープンデータを活用**して、個人の税金の一日あたりの支出額を示しています。
  - 「税金はどこへ行った？」はイギリスの Open Knowledge Foundation の「Where Does My Money Go?」をベースに開発され、Open Knowledge Foundation Japanが運営をしています。

税金はどこへ行った

【出所】税金はどこへ行った［OKFJ］http://spending.jp/

- ウェブアプリケーションSabota（さぼた）では、鯖江市の図書館座席に取り付けた**センサーが発信する情報をオープンデータ**として、空き席情報を表示します。
  - 鯖江市役所では、市民協働推進プロジェクトとして、女子高生が市政に協力する「鯖江市役所JK課」を2014年に発足し、オープンデータを活用するSabotaも女子高生の提案、協力によって構築されました。

Sabota（さぼた）

【出所】Sabota［鯖江市］ http://sabae-jk.jp/app/

□ 自治体全体の支出状況から図書館の各席の空き状況まで、「オープンデータ」は様々なレベルで活用されています。

官民における「オープンデータ」の普及の取り組み

◆官民それぞれの組織で、「オープンデータ」の普及・活用を推進しています。

- 内閣官房の**情報通信技術（IT）総合戦略室**では、**「オープンデータ100」**として、利活用事例を紹介することで、**オープンデータの普及・啓発**を行っています。

オープンデータ100

【出所】オープンデータ100［内閣官房 情報通信技術総合戦略室］
https://cio.go.jp/opendata100

- 民間事業者や自治体が会員となっている一般社団法人**オープン＆ビッグデータ活用・地方創生推進機構（VLED）**では、**オープンデータに関する情報、催しを紹介**しています。

オープン＆ビッグデータ活用・地方創生推進機構

【出所】一般社団法人 オープン＆ビッグデータ活用・地方創生推進機構（VLED）https://www.vled.or.jp/

- 日本国内では、国、民間事業者、自治体と様々な組織がオープンデータの普及・活用を推進しています。

□ 様々な組織が普及・活用を推進している「オープンデータ」をご存知ですか？

「オープンデータ基本指針」におけるオープンデータの定義

◆「オープンデータ基本方針」では、オープンデータの定義として3つの要件を挙げています。

- 内閣官房に設置された高度情報通信ネットワーク社会推進戦略本部（IT総合戦略本部）は、2016年5月に**「新たなオープンデータの展開に向けて」**を公表しました。
  - 「新たな」という言葉が入っているように、それ以前は政府部門の公共データを念頭においていたオープンデータ関連政策を改めて、「オープンデータ 2.0」とし、「民間部門を包含すること」や「課題解決型のオープンデータ政策の推進」へと方針転換が示されました。

【出所】新たなオープンデータの展開に向けて［内閣官房 情報通信技術総合戦略室］https://cio.go.jp/node/1374

- IT総合戦略本部は、2017年5月に国、地方公共団体、事業者が公共データの公開及び活用に取り組む上での基本指針を定めた**「オープンデータ基本方針」**を公表しました。

【出所】オープンデータ基本指針［内閣官房 情報通信技術総合戦略室］https://cio.go.jp/node/2357

- 「オープンデータ基本方針」では、下記の**3つの項目全てに該当するものがオープンデータと定義**しています。

3つ全ての条件を満たすデータがオープンデータ

(1) 営利目的、非営利目的を問わず二次利用可能なルールが適用されたもの

(2) 機械判読に適したもの

(3) 無償で利用できるもの ¥0

- (2)の「機械判読」は、オープンデータ基本方針の脚注において「コンピュータプログラムが自動的にデータを加工、編集等できること」を指すと示されています。「判読」という言葉ではありつつも、コンピュータープログラムによって読めるだけではなく、加工・編集ができることを意味しています。

- 「オープンデータ基本方針」には、『**「機械判読に適した度合」には、人手をどれだけ要さずに、コンピュータがデータを再利用できるかにより、いくつかの段階がある。**』と記載されています。

オープンライセンスの重要性

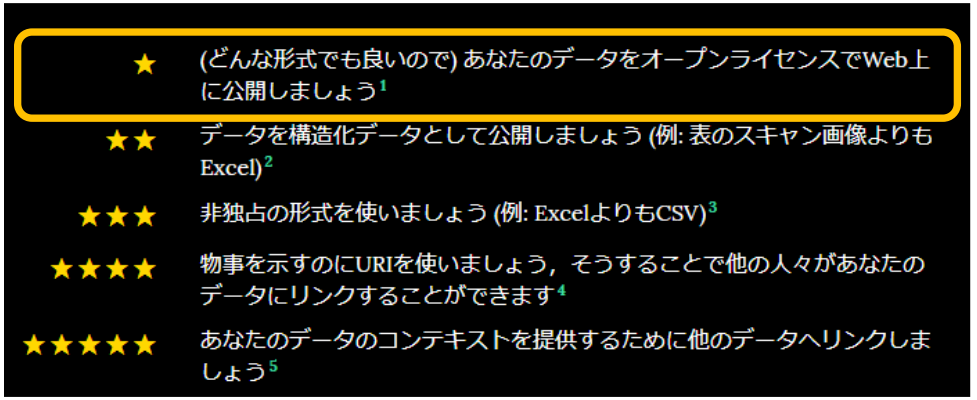
◆「5★オープンデータ」では、オープンライセンスの重要性と機械判読の段階を示しています。

- 総務省のウェブサイトやその出所となった**「5★オープンデータ」**では「オープンデータの5つの段階とデータ形式」として、**オープンデータにおける機械判読のレベル**を示しています。

【出所】オープンデータの意義・目的［総務省］
http://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/opendata/opendata01.html

- 「5★オープンデータ」の顔となっているティム・バーナーズ＝リー博士は「Webの父」とも呼ばれる計算機科学者です。
  - ティム・バーナーズ＝リー博士は1989年にスイスにあるCERN（セルン：欧州原子核研究機構）にてWeb、URL、HTTP、HTMLを設計しました。

「5★オープンデータ」におけるオープンデータの段階に関する紹介



【出所】5★オープンデータ　https://5stardata.info/ja/

- ティム・バーナーズ＝リー博士はオープンデータに加えて、データ同士がつながっているリンクドデータ（Linked Data）の普及に情熱を持っています。博士の意見・情熱は、TEDが無料で提供する動画の「ティム・バーナーズ＝リーが示す次のウェブ」を参照してください。（https://www.ted.com/talks/tim_berners_lee_on_the_next_web/transcript?language=ja）

- 「5★オープンデータ」では、まず**第1段階目としての「オープンライセンスでのWeb公開」**を推奨しています。
  - ここでの「オープンライセンス」とは、自由に利用・他人と共有できるデータを指しており、本講座で後述するCC0をオープンライセンスの例として挙げています。

機械判読への適性の5段階

◆「機械判読への適性」は段階的な指標であり、5段階の指標が示されています。

オープンデータの５つのレベルとデータ形式

	段階	名称	概要	フォーマットの例
データ単体の特性	1段階目	**OL：Open License**	オープンライセンス	PDF、画像データ（jpegなど）
	2段階目	**RE：Reusable**	編集・改訂が可能	Word, Excel, PowerPoint
	3段階目	**OF：Open Format**	ソフトウェアを問わないフォーマット	CSV
外部との関係	4段階目	**URI：Uniform Resource Identifier**	外部からリンクが可能	RDF/XML
	5段階目	**LD：Linked Data**	外部へのリンクがある	外部へのリンクがあるRDF/XML

- PDFや画像データ（jpeg等）は、**自由な編集が困難であるため5段階の1段階目**に位置づけられています。
- Excelファイルのような特定のソフトウェアを使えば、**編集可能なファイルは5段階の第2段階**となっています。
- テキストデータコンマで区切ったCSV（Comma-Separated Values）は、**様々なソフトウェアで編集可能であるため、5段階の第3段階**に位置づけられています。
- **4段階目の「URI」は、外部データから特定やリンクされることが可能**かの基準です。
  - RDF（Resource Description Framework）というデータ・各項目の意味を記述し、外部から特定やリンクされることが可能とする枠組みがあります。
  - 講座1-5でも紹介したファイル形式のXMLには、外部データから特定可能な項目を設定した「RDF/XML」という規格があります。

  【出所】RDF 1.1 XML Syntax［World Wide Web Consortium］ https://www.w3.org/TR/rdf-syntax-grammar/
- **5段階目の「LD」は外部データへのリンク**があり、当該データから外部へつながっているかの基準です。
  - 4段階目までの条件を満たしたファイルに規定に沿った外部へのリンクを設定すれば、5段階目になります。
  - 外部データから特定・リンク可能かつ外部データへのリンクがあり、オープンライセンスで公開されているデータはLOD（Linked Open Data）となります。

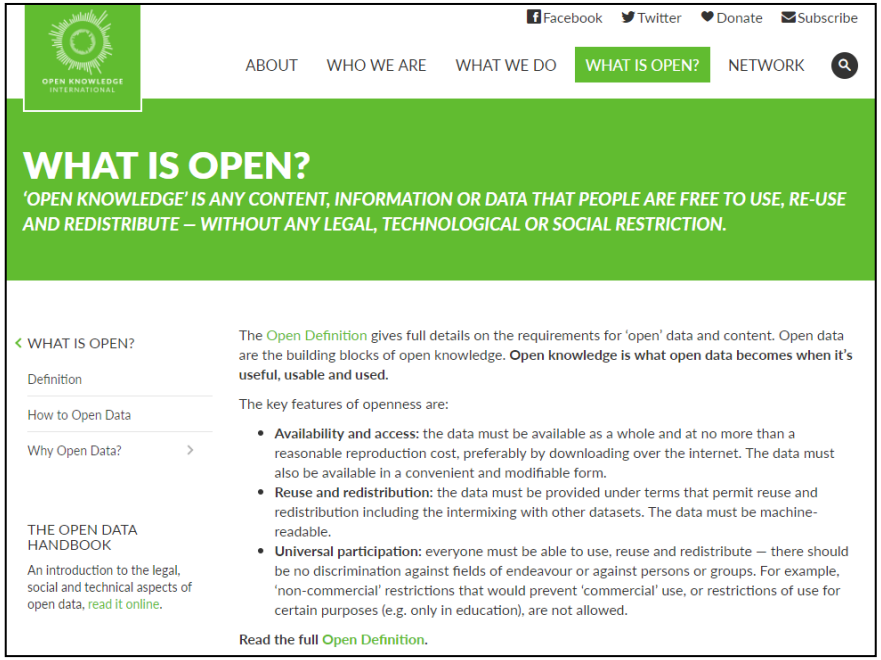
  【出所】Linked Data［World Wide Web Consortium］ https://www.w3.org/DesignIssues/LinkedData.html

OKIにおけるオープンデータの取り組み

◆オープンデータの取り組みは日本のみならず、世界的な広がりを持っています。

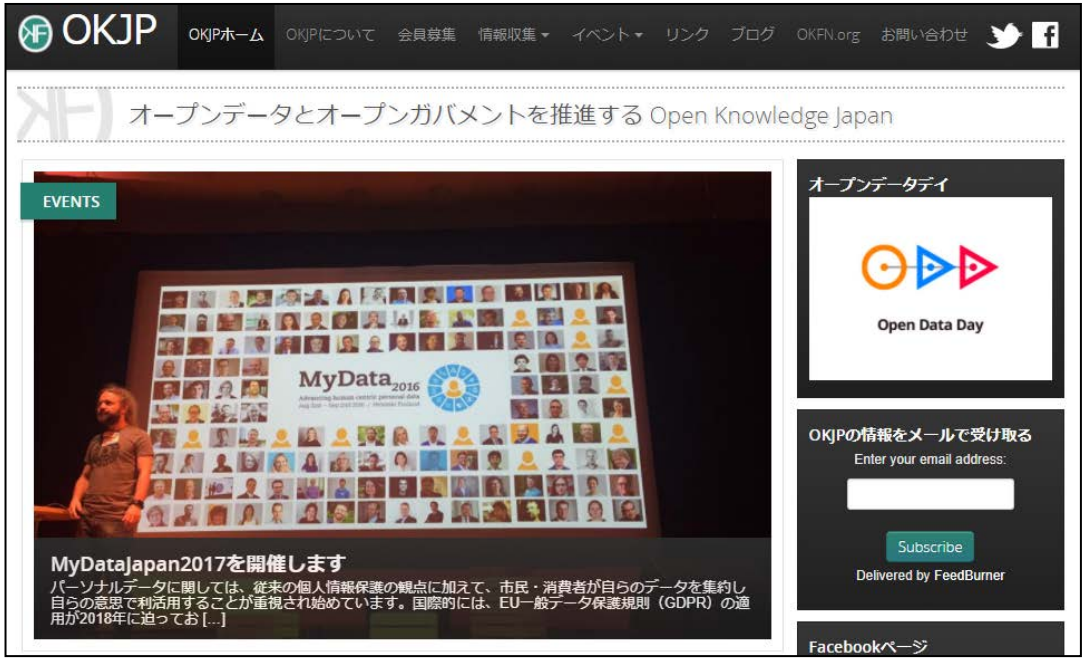
- **Open Knowledge International(OKI)**は、オープンデータの価値を社会に開放することで、社会課題を解決を目指す国際的なNPOネットワークです。
- OKIは40カ国以上の国際的なネットワークを持ち、日本においてはOpen Knowledge Foundation Japanが加盟組織となっています。

Open Knowledge International(OKI)



【出所】Open Knowledge International(OKI)
https://okfn.org/opendata/

Open Knowledge Foundation Japan



【出所】Open Knowledge Foundation Japan
http://okfn.jp/

□ オープンデータの重要性は世界的にも認知されており、国際的な取り組みも進んでいます。

OKIにおけるオープンの定義

◆OKIにおいても、3つの主要な要件を定めてオープンの定義がされています。

● **OKIは「オープン」を「誰でも、どのような目的であっても、自由に利用、改変、再配布できるもの」**と定義しています。より詳細な定義は下記の表の条件を満たすものを「オープン」と呼んでいます。

Open Knowledge Internationalにおけるオープンの定義

主要な特徴	説明
利用、アクセスできる （Availability and access）	データ全体が複製のための適正価格で利用可能であり、インターネットでの無償ダウンロードができることが望ましい。また、データは改変可能な形式でなければならない。
再利用、再配布ができる （Reuse and redistribution）	データは再利用と再配布が可能な形式で提供され、他のデータと統合した再配布が可能であること。また、機械判読が可能であること。
誰でも使える （Universal participation）	分野、個人、グループで差別せず、誰にでも、利用可能、再配布可能であること。 （例えば、「教育目的にのみ利用可能」と差別をしてはならない）

【出所】WHAT IS OPEN? ［Open Knowledge International］ https://okfn.org/opendata/

● 日本政府の「オープンデータ基本方針」に挙げられた「二次利用可能なルール」、「機械判読への適性」、「無償利用」に対して、OKIの定義は観点や強調点は異なるものの方針は類似しています。

- 日本の「オープンデータ基本方針」が無償利用を強調している一方で、OKIでは改変した再配布できることを強調するなど、強調点やニュアンスが異なる部分もあります。
- 2016年にオミダイア・ネットワークのGovLABより公表された「OPEN DATA IMPACT WHEN DEMAND AND SUPPLY MEET」では、オープンデータに関して「There exists no single, universally accepted definition.（唯一かつ普遍的に受けいれられている定義はありません。）」と記し、国際組織や各国政府による異なる10種のオープンデータの定義を示しています。

【出所】OPEN DATA IMPACT WHEN DEMAND AND SUPPLY MEET［Omidyar Network］
http://odimpact.org/files/open-data-impact-key-findings.pdf

OKIにおけるオープンの対象

◆OKIの直近の定義では、オープン「データ」とは言わずに「Open Works」と記述しています。

- OKIでは**Worksの定義として「知識が転換されたもの」**として定義し、**電子ファイルの広い「オープン化」**を推進しています。

Open の定義

Open Definition 2.1

Version 2.1

The Open Definition makes precise the meaning of "open" with respect to knowledge, promoting a robust commons in which anyone may participate, and interoperability is maximized.

Summary: *Knowledge is open if anyone is free to access, use, modify, and share it — subject, at most, to measures that preserve provenance and openness.*

The term **work** will be used to denote the item or piece of knowledge being transferred.

【出所】Open Definition 2.1［Open Knowledge International］
https://opendefinition.org/od/2.1/en/

1. Open Works

An open **work** *must* satisfy the following requirements in its distribution:

1.1 Open License or Status

The **work** *must* be in the **public domain** or provided under an open **license** (as defined in Section 2). Any additional terms accompanying the work (such as a terms of use, or patents held by the licensor) *must not* contradict the work's public domain status or terms of the license.

1.2 Access

The **work** *must* be provided as a whole and at no more than a reasonable one-time reproduction cost, and *should* be downloadable via the Internet without charge. Any additional information necessary for license compliance (such as names of contributors required for compliance with attribution requirements) *must* also accompany the work.

1.3 Machine Readability

The **work** *must* be provided in a form readily processable by a computer and where the individual elements of the work can be easily accessed and modified.

1.4 Open Format

The **work** *must* be provided in an open format. An open format is one which places no restrictions, monetary or otherwise, upon its use and can be fully processed with at least one free/libre/open-source software tool.

□ この背景の一つとして「データ」という言葉が「数値データ」や「構造化データ」を連想させることが挙げられます。

オープンデータにおける「データ」の意味

◆オープンデータにおける「データ」は様々なファイル形式を含みます。

- オープンデータにおける「データ」とは構造化データの数値情報のみならず文書、画像、音楽、動画を含んでおり、様々な形式の電子ファイルを含んでいます。
- 日本語で「オープンデータ」を強引に訳す場合、**「公表データ」**または**「公開データ」**と翻訳するケースが多くなっています。
  しかし、本来のオープンデータは
  - ◆ 単に公開するだけではなく、二次利用を許可する点
  - ◆ 数値データに限らず、様々な電子ファイルを含む点

  から、これらの翻訳は誤解を招きやすい表現となっています。
- Open Knowledge Foundation Japanの代表理事は、「**オープンデータは『ライセンスを開放すること』『様々なファイル形式を含むこと』**から、中国や台湾で使われる**「開放資料」**という訳がふさわしい」という趣旨の意見を表明しています。

【出所】オープンデータが実現するデータ循環社会［worksight］
https://www.worksight.jp/issues/395.html

□ オープンデータの定義は、オープンデータ関連組織によって多少の相異がありますが、広範なファイル形式を含むことは、共通しています。

様々な情報に関するオープンデータ

◆オープンデータはファイル形式のみならず、対象とする情報も多種多様です。

● オープンデータの対象となる情報は、下表のように日常生活に関するデータを含めて多種多様です。

オープンデータの情報の対象例

情報の対象	情報の内容例
地理・地域交通データ	地図・土地利用および公共交通の時刻表・交通量・旅行者数
人口データ	住民の年齢・性別情報および出生・死亡統計
選挙・地方議会のデータ	選挙の投票率・選好、地方議会における議案
予算・税金データ	自治体の予算、税収、支出
犯罪・防犯データ	犯罪統計・警察による検挙率
産業および地域活動データ	産業の状況、地域サービスの実施位置や連絡先
不動産データ	地価や家賃の情報、特定の地域から公共サービスへの移動時間
エネルギーの生産・消費データ	市民・公共団体・産業におけるエネルギーの生産と消費
環境および汚染に関するデータ	大気・水および土壌の汚染率
廃棄物および水管理データ	ごみの廃棄量・収集スケジュール、水の消費量
健康関連データ	病床数、病院のパフォーマンス、病気の感染情報
教育データ	学校の所在地と費用、生徒の性別・年齢・成績の構成
農業・漁業データ	農作物別収穫量、漁獲数
科学に関するデータ	大学、公的研究機関の研究、特許情報等
文化資料	博物館・美術館・展示会・祭り・図書館の資料

【出所】Sashinskaya, Maria. Open Data: All You Want To Know About Open Data (Big Data, Transparency, Urbanism, Transportation, Sustainable Cities, Innovations, Smart Governance, e-government) (p.17)を翻訳して一部改訂

オープンデータに関するウェブサイト

◆国、自治体を中心に様々なオープンデータ関連サイトを運営しています。

● 国は、様々なオープンデータを閲覧・ダウンロードできるポータルサイトを運営しています。

- Data.GO.JPでは、データを公開することの重要性から、PDF形式を含めて国が公表した様々な資料を公開しています。
- 公的統計の結果の公表値はオープンデータの一種であり、公的統計のポータルサイトであるe-Statから、検索・ダウンロードができます。

DATA.GO.JP（オープンデータカタログサイト）

【出所】DATA.GO.JP［総務省］http://www.data.go.jp/

e-Stat（政府統計の総合窓口）

【出所】e-Stat［総務省］https://www.e-stat.go.jp/

● 各自治体においても、様々なオープンデータに関するウェブサイトの公開が進展しています。

- 様々なオープンデータを網羅した「東京都オープンデータカタログサイト」もあれば、道路情報に特化した「しずみち info」もあります。

東京都オープンデータカタログサイト

【出所】東京都オープンデータカタログサイト［東京都］
http://opendata-portal.metro.tokyo.jp/www/

しずみちinfo

【出所】しずみちinfo［静岡市］
https://shizuokashi-road.appspot.com/

画像形式のオープンデータを集めたウェブサイト

◆画像形式のオープンデータを専門に収集、公開しているウェブサイトがあります。

- 神戸市が提供する［阪神・淡路大震災「1.17の記録」］では、震災時や復旧に至る写真がオープンデータとして提供されています。
  - 「1.17の記録」は、出所表記さえすれば、営利目的でも使える写真であり、続いて説明するクリエイティブ・コモンズ（CC BY）に該当します。

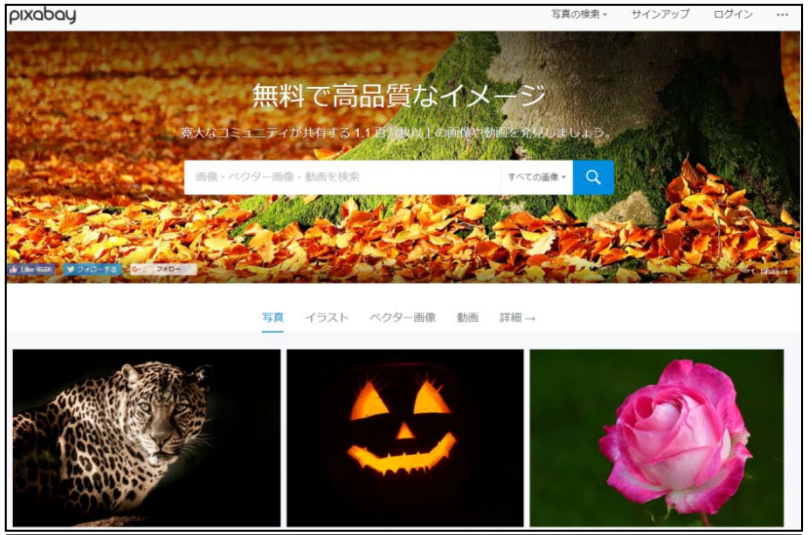
阪神・淡路大震災「1.17の記録」

写真の利用にあたり、神戸市に申請や連絡が必要ですか？
いずれも不要です（一部注意書きのある写真を除きます）。
「写真提供：神戸市」等のクレジットを表示して自由にご利用ください。

【出所】阪神・淡路大震災「1.17の記録」［神戸市］
http://kobe117shinsai.jp/

- 無料画像の提供サイトのpixabayでは、出所表記不要で利用できる画像を検索することができます。
  - pixabayで提供されているデータは、続いて説明する「クリエイティブ・コモンズCC0」として、出所記載も不要で自由な利用ができるオープンデータです。

無料画像の提供サイトのpixabay



すべてのコンテンツはクリエイティブ・コモンズCC0の下で公開されています。これにより、たとえ商業目的であっても、許可を求めたりアーティストに信用を与えることなく安全に使用できます。

【出所】pixabay
https://pixabay.com/

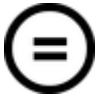
□ 続いて、オープンデータに関連する事項として「クリエイティブ・コモンズ」に関して説明します。

クリエイティブ・コモンズにおける4種の条件

◆オープンデータに関連して、4種の条件で構成されるクリエイティブ・コモンズを紹介します。

- **クリエイティブ・コモンズ**のライセンスは、作品の公開者が**「この条件を守れば、私の作品を自由に使って構いません。」という意思表示をする国際的な許諾の枠組み**です。
  - クリエイティブ・コモンズ（Creative Commons）はCCと略され、クリエイティブ・コモンズのライセンスはCCライセンスとも表記されます。
  - クリエイティブ・コモンズ・ジャパンのFAQには「CCライセンスは、日本の著作権法その他の法律に基づいていますので、原則としては法律的な拘束力があります。」と記載されています。
- **CCライセンス**を利用することで、**作者は著作権を保持したまま作品を自由に流通**させることができ、**作品の利用者はライセンス条件の範囲内で改変や再配布**をすることができます。
- クリエイティブ・コモンズは、下記の**4種の条件で構成**されています。
  - CCライセンスにおいて、「営利目的の利用禁止」「改変禁止」が条件に入る場合は、日本政府やOKIが示す”オープン”の定義から外れます。

クリエイティブ・コモンズにおける4種の条件

アイコン	英語表記	日本語訳	説明
	BY （Attribution）	**表示**	作品の出所・クレジットを表示すること ［タイトル、全ての著作者、URLを表示］
	NC （Non-commercial）	**非営利**	営利目的での利用をしないこと
	**ND** （No Derivative Works）	**改変禁止**	元の作品を改変しないこと
	SA （Share-alike）	**継承**	元の作品と同じ組み合わせのCCライセンスで公開すること

【出所】クリエイティブ・コモンズ・ライセンスとは［クリエイティブ・コモンズ・ジャパン］を改変して作成
https://creativecommons.jp/licenses/

クリエイティブ・コモンズの6種類のライセンス

◆クリエイティブ・コモンズでは、条件の組み合わせで6種類のライセンスが示されています。

● **標準的なCCライセンスは「表示」は必須**とし、残る**3つの条件の組み合わせで6種類のライセンス**があります。

- 3種の組み合わせ(2^3)は、本来8種ですが、「継承」は「改変が許可されている」ことが前提となるため、6種類のライセンスとなります。
- 2018年3月現在において、2013年11月に公開された英語版、2015年7月に日本語版が正式公開されたバージョン4.0が最新のCCライセンスです。

クリエイティブ・コモンズの6種のライセンスと利用条件

画像	利用ルール名称	利用の許可・不許可		
		出所の表示	商業利用	改変
	表示 4.0 国際 (CC BY 4.0)	表示は必須	商業利用を許可する	改変を許可する
	表示-非営利 4.0 国際 (CC BY-NC 4.0)		商業利用を許可しない（改変されたものを含む）	改変を許可する
	表示-改変禁止 4.0 国際 (CC BY-ND 4.0)		商業利用を許可する	改変を許可しない
	表示-非営利-改変禁止 4.0 国際 (CC BY-NC-ND 4.0)		商業利用を許可しない	改変を許可しない
	表示-継承 4.0 国際 (CC BY-SA 4.0)		商業利用を許可する	改変を許可するが、利用ルールは踏襲
	表示-非営利-継承 4.0 国際 (CC-NC-SA 4.0)		商業利用を許可しない（改変されたものを含む）	改変を許可するが、利用ルールは踏襲

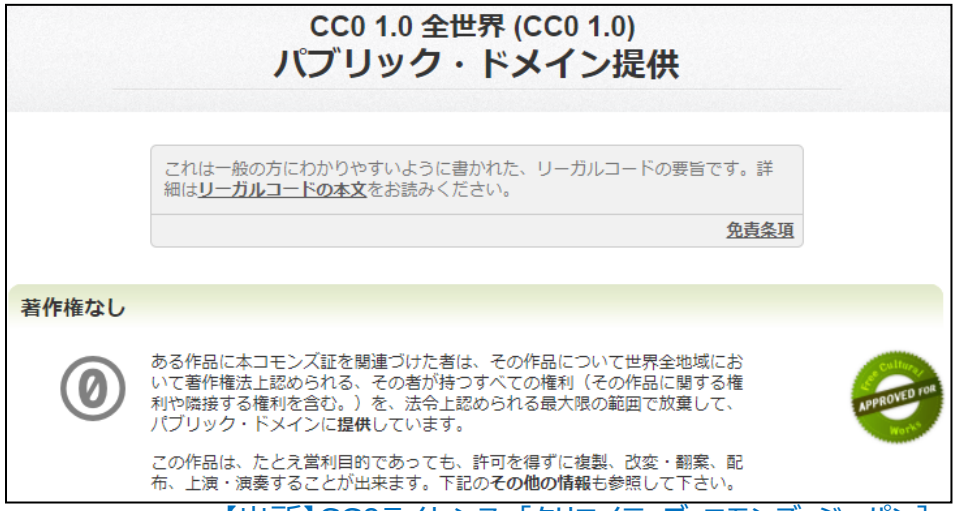
【出所】クリエイティブ・コモンズ・ライセンスとは［クリエイティブ・コモンズ・ジャパン］を改変して作成 https://creativecommons.jp/licenses/

CC0とパブリック・ドメイン（公有、公共の領域）

◆出所表記不要の利用形態として、「CC0（パブリック・ドメイン）」が挙げられます。

- CCライセンスの特殊な形態としての、**CC0（シー・シー・ゼロ）**は、**法令上認められる最大限放棄して、パブリック・ドメイン（公有、公共の領域）に提供**する形態を指しています。
  - CCライセンスの標準的な6つの形態は、いずれも「作者は著作権を保持」したまま作品を流通させるため、「表示」が必須となっていました。
  - CC0であれば、出所（クレジット）の表記も不要で、自由に利用することができます。
- **CC0は、著作権者が意図的に権利を放棄**して、**パブリック・ドメイン**に提供した形態を指しています。
  - 従来のパブリックドメインは、著作権者の死後50年などの保護期間が終了したものを指すことが一般的でした。
- 著作権者の死後50年が経過し、パブリックドメインになった文献を読めるウェブサイトとして「青空文庫」が挙げられます。

CC0 1.0 全世界 (CC0 1.0) パブリック・ドメイン提供



【出所】CC0ライセンス［クリエイティブ・コモンズ・ジャパン］
https://creativecommons.jp/2015/05/01/cc0-jpver/

パブリックドメインになった文献が読める「青空文庫」

【出所】青空文庫
https://www.aozora.gr.jp/

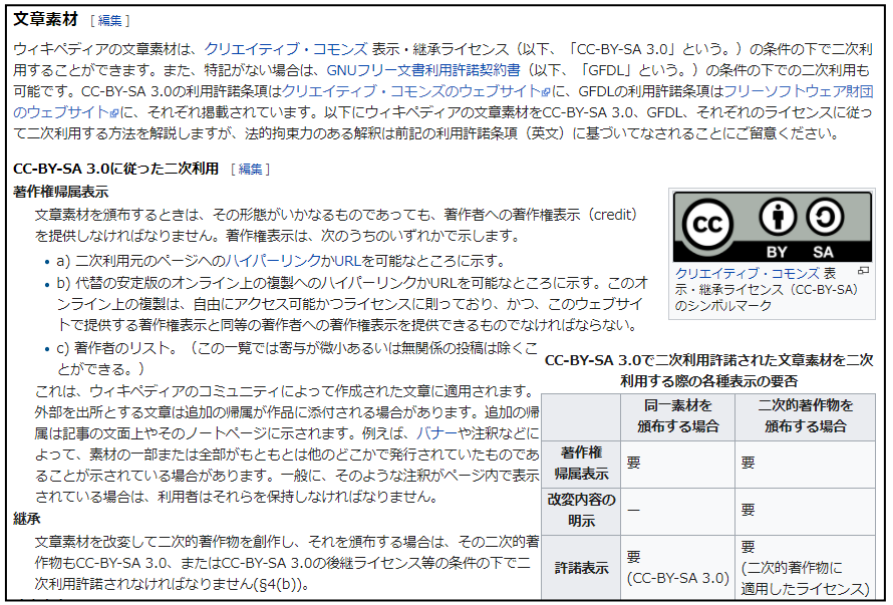
- 標準的なCCライセンスの中でも「営利目的で利用可能」かつ「改変可能」であるケース、CC0をはじめパブリックドメインとなっているケースは、日本政府やOKIが示す「オープン」の条件に該当します。

誰でもオープンデータを投稿・発信できるウェブサイト

◆Wikipediaなど、オープンデータを誰でも投稿・発信できるウェブサイトもあります。

- Wikipediaにおける文章素材は「表示-継承 (CC-BY-SA)」の条件で二次利用可能であり、誰でも投稿・発信できるオープンデータのウェブ百科辞典と言えます。
  - Wikipediaの文章素材は、「営利目的の利用可能」「（改変箇所を明示すれば）改変可能」であることから、日本政府やOKIが示すオープンデータの定義に合致します。

ウィキペディアを二次利用する



文章素材 ［編集］

ウィキペディアの文章素材は、クリエイティブ・コモンズ 表示・継承ライセンス（以下、「CC-BY-SA 3.0」という。）の条件の下で二次利用することができます。また、特記がない場合は、GNUフリー文書利用許諾契約書（以下、「GFDL」という。）の条件の下での二次利用も可能です。CC-BY-SA 3.0の利用許諾条項はクリエイティブ・コモンズのウェブサイトに、GFDLの利用許諾条項はフリーソフトウェア財団のウェブサイトに、それぞれ掲載されています。以下にウィキペディアの文章素材をCC-BY-SA 3.0、GFDL、それぞれのライセンスに従って二次利用する方法を解説しますが、法的拘束力のある解釈は前記の利用許諾条項（英文）に基づいてなされることにご留意ください。

CC-BY-SA 3.0に従った二次利用 ［編集］

著作権帰属表示

文章素材を頒布するときは、その形態がいかなるものであっても、著作者への著作権表示（credit）を提供しなければなりません。著作権表示は、次のうちのいずれかで示します。

- a) 二次利用元のページへのハイパーリンクかURLを可能なところに示す。
- b) 代替の安定版のオンライン上の複製へのハイパーリンクかURLを可能なところに示す。このオンライン上の複製は、自由にアクセス可能かつライセンスに則っており、かつ、このウェブサイトで提供する著作権表示と同等の著作者への著作権表示を提供できるものでなければならない。
- c) 著作者のリスト。（この一覧では寄与が微小あるいは無関係の投稿は除くことができる。）

これは、ウィキペディアのコミュニティによって作成された文章に適用されます。外部を出所とする文章は追加の帰属が作品に添付される場合があります。追加の帰属は記事の文面上やそのノートページに示されます。例えば、バナーや注釈などによって、素材の一部または全部がもともとは他のどこかで発行されていたものであることが示されている場合があります。一般に、そのような注釈がページ内で表示されている場合は、利用者はそれらを保持しなければなりません。

継承

文章素材を改変して二次的著作物を創作し、それを頒布する場合は、その二次的著作物もCC-BY-SA 3.0、またはCC-BY-SA 3.0の後継ライセンス等の条件の下で二次利用許諾されなければなりません(§4(b))。

クリエイティブ・コモンズ 表示・継承ライセンス（CC-BY-SA）のシンボルマーク

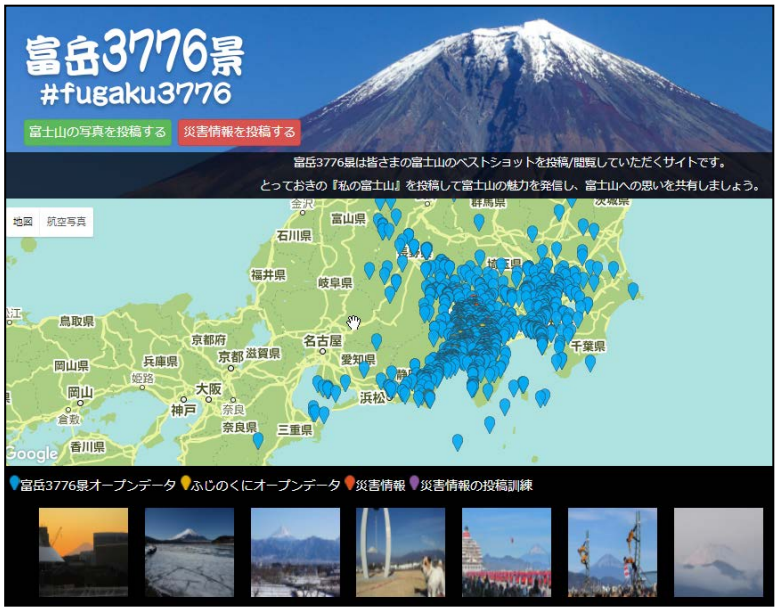
CC-BY-SA 3.0で二次利用許諾された文章素材を二次利用する際の各種表示の要否

	同一素材を頒布する場合	二次的著作物を頒布する場合
著作権帰属表示	要	要
改変内容の明示	—	要
許諾表示	要 (CC-BY-SA 3.0)	要 (二次的著作物に適用したライセンス)

【出所】ウィキペディアを二次利用する［Wikipedia 日本語版］
https://ja.wikipedia.org/w/index.php?title=WP:REUSE

- 様々な場所からの富士山の写真を収集している富岳3776景は、投稿された写真を「表示(CC-BY)」のオープンデータとして扱っています。
  - 富岳3776景に投稿された写真は、「出所：富岳3776景」またはURLの表記をすることで、「営利目的の利用可能」「改変可能」であり、オープンデータと見なすことができます。

富岳3776景



【出所】富岳3776景［静岡県・山梨県］
http://fugaku3776.okfn.jp/

□ これらのウェブサイトでは、誰でもオープンデータを投稿することができ、投稿されたデータを誰もが利用できます。

相互運用性（interoperability：インターオペラビリティ）

◆OKIでは”オープン”の重要性に、「相互運用性（interoperability）」を挙げています。

【英語原文】If you’re wondering why it is so important to be clear about what open means and why this definition is used, there’s a simple answer: **interoperability**.

【OKIによる日本語訳】なぜそこまでして「オープン」の意味をはっきりさせたいの？ なぜその定義を使うことにしたの？ その答えは、一言で表すと **相互運用性（インターオペラビリティ）** だ。

- IT用語事典［株式会社インセプト］では「インターオペラビリティとは、複数の異なるものを接続したり組み合わせて使用したときに、きちんと全体として正しく動作すること。また、その度合い」と記しています。

【OKIによる日本語訳】**相互運用性が大切な理由は、さまざまなコンポーネントを組み合わせて使えるようになる**からだ。コンポーネントにまとめてそれらを**「つなぎ合わせて」使う**という仕組みは、大規模で複雑なシステムを構築するときに欠かせない。相互運用性がなければ、これは事実上不可能だ。かの有名なバベルの塔の神話が示すとおり、お互いにコミュニケーションができなければ塔を建築する作業は完全に崩壊してしまう。

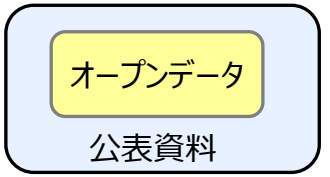
【OKIによる日本語訳】オープンであることについての明確な定義を用意すれば、別々の場所から二つのオープンデータセットを入手したときにそれを組み合わせて使えるようになる。そして **「バベルの塔」状態、つまり大量のデータセットがあるのにそれを組み合わせて(真の価値を持つ)大規模なシステムを作れない状態になってしまうことを回避できる。**

【出所】OPEN DATA HANDBOOK［Open Knowledge International］
http://opendatahandbook.org/guide/en/what-is-open-data/

- オープンデータ関連組織ごとに「オープン」の定義は多少異なりますが、「相互運用性（インターオペラビリティ）を発揮できるデータであるべき」という点は、どの組織の定義においても共通しています。

公表資料に対する引用

◆公表資料であれば引用はできますが、オープンデータでなければ活用ができません。



- オープンデータでなくても、公表された資料であれば、**引用**することはできます。
  - 引用は著作権法第32条で認められた合法的な行為であり、権利者に無断で行われることが一般的です。
  - 公表された資料とは、販売している書籍や上映されている映画など、無償利用ができないものを含みます。
- 著作権法における**正当な引用**であるためには、次の**4つの要件を満たす必要**があります。

著作権法における引用の4つの要件

法	引用の要件	補足説明
32条	**引用する資料等は既に公表されているものであること**	著作権法における公表の定義は、第4条に記載されていますが、例えば『書籍の形で市販、一般向けに上映・演奏、インターネット上で一般公開』していれば、公表していると見なすことができます。
	「公正な慣行」に合致すること	例えば、『引用を行う「必然性」があることや、言語の著作物についてはカギ括弧などにより「引用部分」が明確になってくること』などが挙げられます。
	報道、批評、研究などの引用の目的上「正当な範囲内」であること	例えば、『引用部分とそれ以外の部分の「主従関係」が明確であることや、引用される分量が必要最小限度の範囲内であること』などが挙げられます。
48条	**出所の明示が必要なこと**	引用における出所の表示を確認した者が、出所の著作物にアクセスできる程度に特定する必要があります。

【出所】著作権なるほど質問箱［文化庁］に基づき作成　https://pf.bunka.go.jp/chosaku/chosakuken/naruhodo/ref.asp

- オープンデータでなくてもルールに沿った引用はできますが、権利者の許諾なく著作物を活用することができません。
- **オープンデータには相互運用性があるため、権利者に断りなく改変し、つなぎ合わせて活用**することができます。

相互運用性になって可能となる分析・サービス

◆オープンデータの相互運用性によって、データを組み合わせたサービスが可能になります。

- 相互運用性があるデータであれば、同種のデータを集めて**規模（スケール）を**拡大して利用したり、異なる種類のデータを組み合わせて**範囲（スコープ）を拡大して利用**することができます。
  - 講座1-5でも紹介したように、複数種のデータを加工したり、統合したりすることで、新たなウェブサービスを生み出すことを「マッシュアップ」といいます。

規模（スケール）拡大の事例

- 「AEDオープンデータ プラットフォーム」では、複数の地域のデータを統合して広域データを提供しています。

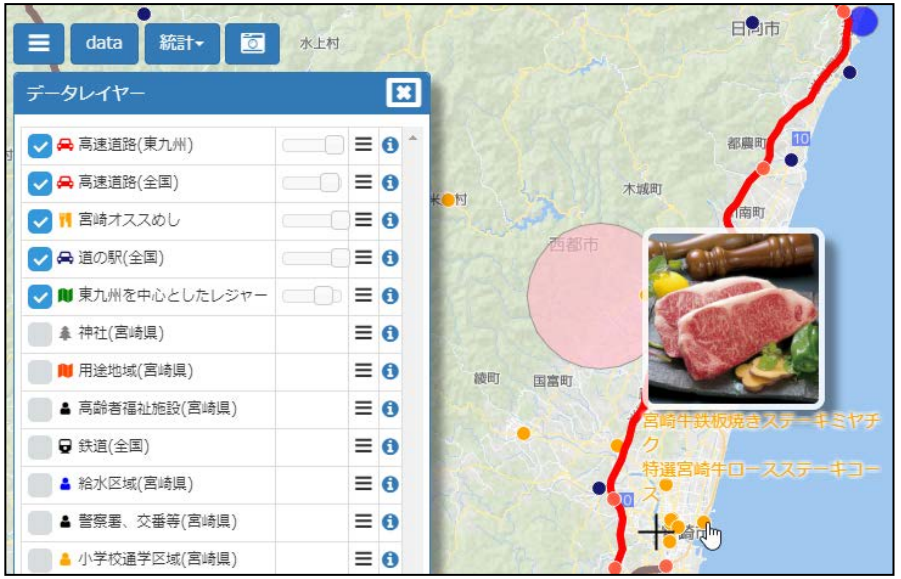
AEDオープンデータ プラットフォーム

【出所】AED オープンデータプラットフォーム
http://hatsunejournal.jp/w8/AEDOpendata/

範囲（スコープ）拡大の事例

- ひなたGISでは、オープンデータを統合して地図、交通網、気象、飲食店といった情報を表示しています。

ひなたGIS（地理情報システム）



【出所】ひなた GIS［宮崎県］
https://hgis.pref.miyazaki.lg.jp/hinata/

□ 講座1-5にて示したAPIを通じて、オープンデータが提供されることで、種々の複合サービスが可能になります。

オープンデータビジネスの3つのタイプ

◆オープンデータを利用するビジネスは、3つのタイプに分けることができるとされています。

● **オープン&ビッグデータ活用・地方創生推進機構（VLED）**が公開している**「オープンデータ利活用ビジネス事例集」によれば**、オープンデータを活用したビジネスは下表の**付加価値型**、**新価値創造型**、**プラットフォーム型**の3種に大別できるとしています。

オープンデータビジネスの3つのタイプ

タイプ	特徴	主な担い手企業
付加価値型	• **既存ビジネスの価値を高めるためにオープンデータを利用する** • データの加工は可視化などが主であり、複雑な処理はしない • 競合相手もオープンデータを自由に利用できるため、既存ビジネスの優劣を極端に変えることはない	市場のリーダー
新価値創造型	• **オープンデータを含む多様なデータをかけ合わせ、高度な分析によって未来を予測する** • 価値を生み出す源泉は新しく開発したアルゴリズムや分析モデル • オープンデータはアルゴリズムや分析モデルを開発する際にも利用される	スタートアップ
プラットフォーム型	• **特定の領域のデータを大量に集め、プラットフォーム化する** • 集めたデータを利用しやすく提供することで最初の価値を生み出す • データの利用状況や利用者の状況を分析することで、さらに新しい価値を生み出していく	スタートアップ

【出所】オープンデータ利活用ビジネス事例集［オープン&ビッグデータ活用・地方創生推進機構］
http://www.vled.or.jp/results/opendata_business_usecases.pdf

オープンデータビジネスの事例

◆3種のオープンデータビジネスの事例をそれぞれ示します。

付加価値型の事例

● ナビタイムジャパンでは「各自治体が公開する避難場所情報」をオープンデータとして利用し、地図表示やルート検索ができるようにしています。

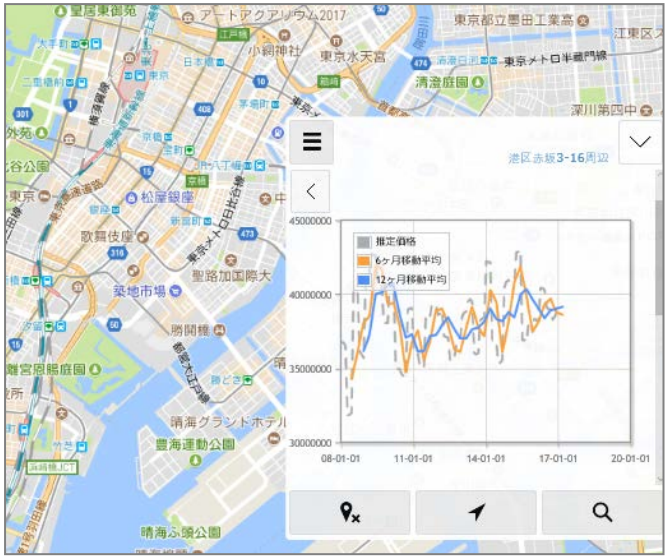
ナビタイムジャパンの避難所の表示

【出所】NAVITIME［株式会社ナビタイムジャパン］
https://www.navitime.co.jp/

新価値創造型 の事例

● GEEOでは「人口や住宅・土地に関する公的統計」をオープンデータとして活用して、不動産販売価格の予測サービスを提供しています。

GEEOの予測不動産価格の表示



【出所】GEEO［株式会社おたに］
https://geeo.otani.co/JPN/

プラットフォーム型の事例

● マイ広報誌では「自治体の広報誌」をオープンデータとして活用して、インターネット配信サービスを提供しています。

地域の広報誌を読めるマイ広報誌

【出所】マイ広報誌［オープン・コーポレイツ・ジャパン］https://mykoho.jp/

□ オープンデータの公開や利活用を推進することで、新たなビジネスが創出されています。

総務省　ICTスキル総合習得教材

［コース4］オープンデータ・ビッグデータ利活用事例

4-2：地域活性化に向けたデータの利活用（RESAS）

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_4_2.pdf

	1	2	3	4	5
［コース1］データ収集					
［コース2］データ蓄積					
［コース3］データ分析					
［コース4］データ利活用		●			

本講座の学習内容［4-2：地域活性化に向けたデータの利活用（RESAS）］

【講座概要】

- 内閣官房（まち・ひと・しごと創生本部事務局）が運営している地域情報を可視化するWebアプリケーションの地域経済分析システム（RESAS）を説明します。
- RESASの構築背景や出所となっているオープンデータ・ビッグデータの概況を紹介します。
- RESASに関する学習機会や普及の取り組みを説明します。
- 分析事例を示しながら、RESASの基本的な操作方法を紹介します。

【講座構成】

座学	[1] RESASとオープンデータ・ビッグデータ
実習	[2] 人口マップの活用
	[3] 産業構造マップの活用
	[4] 観光マップ・まちづくりマップの活用
	[5] 地方財政マップの活用

【学習のゴール】

- ✓ 地域経済分析システム（RESAS）の構築背景や出所となっているデータの概況を理解する。
- ✓ RESASに関する学習機会や普及の取り組みを理解する。
- ✓ RESASの基本操作を理解し、RESASによる分析を行えるようになる。

RESAS（地域経済分析システム）

◆“RESAS（リーサス）”とも呼ばれる「地域経済分析システム」は、オープンデータ・ビッグデータに基づき地域情報を可視化するWebアプリケーションです。

- この講座では**オープンデータ・ビックデータを活用**したWebアプリケーションの**RESAS（地域経済分析システム）**を紹介します。
- RESAS（地域経済分析システム）は経済産業省が開発し、**内閣官房（まち・ひと・しごと創生本部事務局）が運営**を行っているオープンデータ・ビッグデータに基づき**地域情報を可視化するWebアプリケーション**です。
- 正式名称「地域経済分析システム」は、英語表記（**R**egional **E**conomy (and) **S**ociety **A**nalyzing **S**ystem）の頭文字を取って、**通称「RESAS（リーサス）」**と呼ばれています。

RESASトップページ

【出所】RESAS［内閣官房（まち・ひと・しごと創生本部事務局）］https://resas.go.jp/

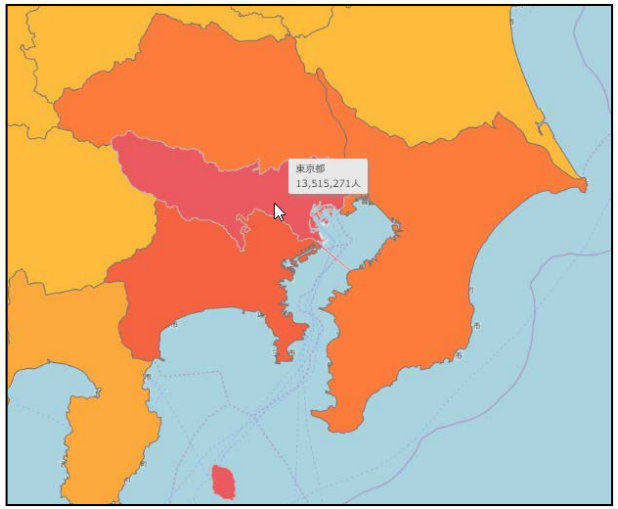
RESASの開発背景

◆RESASは、人口動態、産業構造、観光資源等を可視化できるWebアプリケーションです。

- 人口動態、産業構造、観光資源は地域によって異なるため、**地域を活性化させるために有効な政策・取り組みは地域によって相異**があります。

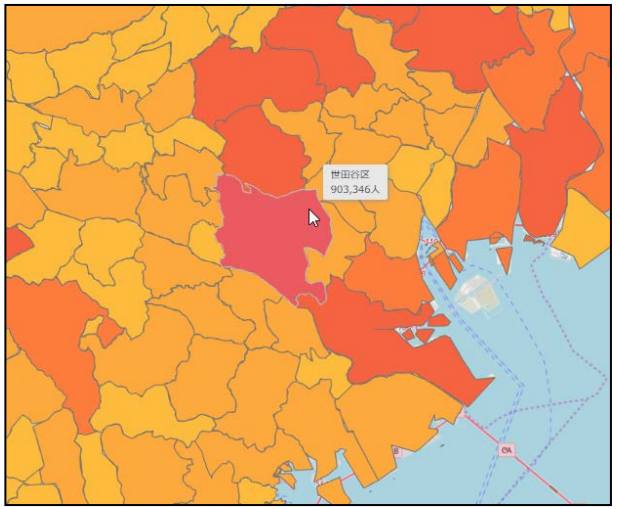
- RESASは各地域の状況を把握、共有するために**地域データの情報基盤**として構築されました。
  - RESASはPCやスマートフォンのWebブラウザ（Google Chrome/Internet Explorer等）を通じて、いつでも、誰でも利用することができます。
  - 出力できる地域単位の情報は都道府県単位のケース、市区町村単位のケース、市区町村内をメッシュでより詳細に表示できるケースがあります。

［人口マップ：人口構成］
における都道府県単位の表示



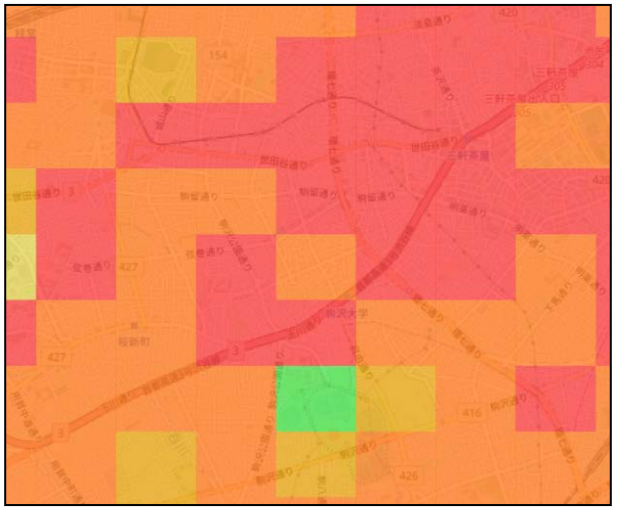
【出所】RESAS（人口マップ：人口構成）

［人口マップ：人口構成］
における市区町村単位の表示



【出所】RESAS（人口マップ：人口構成）

［人口マップ：人口メッシュ］
における市区町村内のメッシュ表示



【出所】RESAS（人口マップ：人口メッシュ）

- RESASによって、地域住民全体で**情報共有し、地域活性化の政策や取り組みを議論・検討**することができます。

RESASとオープンデータ・ビッグデータ

◆RESASはオープンデータ・ビッグデータを可視化するWebアプリケーションです。

- RESASが利用している**オープンデータとして、公的機関が作成したデータ**（公的統計および公的統計以外の官庁作成データ）が挙げられます。
  - 公的統計自体がデータとして表示されるケースもあれば、公的機関による予測値・推計値が表示されるケースもあります。
- RESASでは、GPS等によるモバイル端末の位置情報、スーパーやコンビニエンスストアのレジにて収集されたPOSデータ等の**民間事業者のビッグデータの集計情報を表示**することができます。
- 2018年3月時点において、RESASには**マップと呼ばれる8つの分析メニューの大分類**があり、**合計78の一般メニュー（分析のツール）**があります。
  - RESASには、アカウントを持つ公務員のみが利用できる限定メニューが3つありますが、ほとんどは誰でも利用できる一般メニューです。

RESASのマップ構成と一般メニュー数（2018年3月1日時点）

大分類（マップ）	中分類	一般メニュー数
人口マップ	-	8
地域経済循マップ	-	5
産業構造マップ	全産業/製造業/小売・卸売業（消費）/農業/林業/水産業	28
企業活動マップ	企業情報/海外取引/研究開発	9
観光マップ	国内/外国人	12
まちづくりマップ	-	7
雇用/医療・福祉マップ	-	5
地方財政マップ	-	4

RESASのデータ出所（1/2）

◆RESASの各データの出所（2018年3月現在：限定メニュー除く）

大分類（マップ）	中分類	小分類	出所
人口マップ	—	人口構成	総務省「人口推計」、国立社会保障・人口問題研究所「日本の地域別将来推計人口」
		人口増減	
		人口の自然増減	厚生労働省「人口動態調査」
		人口の社会増減	総務省「住民基本台帳人口移動報告」
		新卒者就職・進学	厚生労働省「雇用動向調査」、文部科学省「学校基本調査」
		将来人口推計	国立社会保障・人口問題研究所「日本の地域別将来推計人口」に基づきまち・ひと・しごと創生本部作成
		人口メッシュ	総務省「国勢調査に関する地域メッシュ統計」
		将来人口メッシュ	国土交通省「メッシュ別将来人口推計」
地域経済循環マップ	—	地域経済循環図	環境省「地域産業連関表」、「地域経済計算」 （株式会社価値総合研究所（日本政策投資銀行グループ）受託作成）
		生産分析	
		分配分析	
		支出分析	
		労働生産性等の動向分析	内閣府「県民経済計算」、総務省「国勢調査」「人口推計」、社人研「将来推計人口」
産業構造マップ	全産業	全産業の構造	総務省「経済センサス－基礎調査」再編加工、総務省・経済産業省「経済センサス－活動調査」再編加工
		稼ぐ力分析	総務省・経済産業省「経済センサス－活動調査」再編加工
		企業数	総務省「経済センサス－基礎調査」、総務省・経済産業省「経済センサス－活動調査」
		事業所数	
		従業者数（事業所単位）	総務省「経済センサス－基礎調査」、総務省・経済産業省「経済センサス－活動調査」再編加工
		付加価値額（企業単位）	総務省・経済産業省「経済センサス－活動調査」再編加工
		労働生産性（企業単位）	
	製造業	製造業の構造	経済産業省「工業統計調査」再編加工、総務省・経済産業省「経済センサス－活動調査」再編加工
		製造業の比較	
		製造品出荷額等	経済産業省「工業統計調査」、総務省・経済産業省「経済センサス－活動調査」
	小売・卸売業（消費）	商業の構造	経済産業省「商業統計調査」
		商業の比較	
		年間商品販売額	
		消費の傾向（POSデータ）	カスタマー・コミュニケーションズ株式会社「TRUE DATA」
		From-to分析（POSデータ）	
	農業	農業の構造	農林水産省「農林業センサス」再編加工
		農産物販売金額	
		農地分析	
		農業者分析	
	林業	林業総収入	
		山林分析	
		林業者分析	
	水産業	海面漁獲物等販売金額	農林水産省「漁業センサス」再編加工
		海面漁船・養殖面積等分析	
		海面漁業者分析	
		内水面漁獲物等販売金額	
		内水面漁船・養殖面積等分析	
		内水面漁業者分析	

RESASのデータ出所（2/2）

大分類（マップ）	中分類	小分類	出所
企業活動マップ	企業情報	表彰・補助金採択	経済産業省「法人インフォメーション」
		創業比率	総務省「事業所・企業統計調査」、総務省「経済センサス－基礎調査」、総務省・経済産業省「経済センサス－活動調査」
		黒字赤字企業比率	総務省・経済産業省「経済センサス－活動調査」再編加工
		中小・小規模企業財務比較	一般社団法人CRD協会
	海外取引	海外への企業進出動向	経済産業省「海外事業活動基本調査」再編加工
		輸出入取引	財務省「貿易統計」
		企業の海外取引額分析	経済産業省「企業活動基本調査」再編加工
	研究開発	研究開発費の比較	経済産業省「企業活動基本調査」再編加工
		特許分布図	特許庁「特許情報」
観光マップ	国内	目的地分析	株式会社ナビタイムジャパン「経路検索条件データ」
		From-to分析（宿泊者）	経済産業省「観光予報プラットフォーム」
		宿泊施設	観光庁「宿泊旅行統計調査」
	外国人	外国人訪問分析	観光庁「訪日外国人消費動向調査」、日本政府観光局（JNTO）「訪日外客数」
		外国人滞在分析	株式会社NTTドコモ・株式会社ドコモ・インサイトマーケティング「モバイル空間統計®」
		外国人メッシュ	
		外国人入出国空港分析	
		外国人移動相関分析	
		外国人消費の比較（クレジットカード）	ビザ・ワールドワイド・ジャパン株式会社のカードデータに基づき、内閣官房まち・ひと・しごと創生本部事務局作成
		外国人消費の構造（クレジットカード）	
		外国人消費の比較（免税取引）	観光庁「消費税免税店（輸出物品販売場）の都道府県別分布」
		外国人消費の構造（免税取引）	グローバルブルー ティエフエス ジャパン株式会社「免税利用状況データ」
まちづくりマップ	―	From-to分析（滞在人口）	株式会社NTTドコモ・株式会社ドコモ・インサイトマーケティング「モバイル空間統計®」
		滞在人口率	
		通勤通学人口	総務省「国勢調査」
		流動人口メッシュ	株式会社Agoop「流動人口データ」
		事業所立地動向	株式会社日本ソフト販売「電話帳データ」
		施設周辺人口	国土交通省「メッシュ別将来人口推計」、「国土数値情報」
		不動産取引	国土交通省「土地総合情報システム　不動産取引価格情報」
雇用/医療・福祉マップ	―	一人当たり賃金	厚生労働省「賃金構造基本統計調査」
		有効求人倍率	厚生労働省「職業安定業務統計」
		求人・求職者	
		医療需給	厚生労働省「医療施設静態調査」、「医師・歯科医師・薬剤師調査」、「衛生行政報告例」、「患者調査」、総務省「人口推計」、「住民基本台帳に基づく人口」
		介護需給	厚生労働省「地域包括ケア『見える化』システム」、「介護サービス施設・事業所調査」
地方財政マップ	―	自治体財政状況の比較	総務省「地方財政状況調査関係資料（財政状況資料集）」「国勢調査」「住民基本台帳」
		一人当たり地方税	総務省「地方財政状況調査」
		一人当たり市町村民税法人分	
		一人当たり固定資産税	

より詳細なデータ説明、用語の解説に関しては、RESASの画面右上の「ヘルプ」から、「操作マニュアル」の「各章末のデータ説明」や「データ出所一覧」を確認してください

RESEASの普及の取り組み

◆RESASには、オンライン講座、API、ウェブコミニティなどの関連ウェブサービスがあります。

- **RESASのオンライン講座**には基礎編と応用編があり、それぞれの講座を終了すると、修了証が発行されます。
  - 基礎編では、RESASに関する知識がない方でも、操作方法や出力の見方から学ぶことができます。
  - 応用編では、RESASの複数のマップやRESAS以外のデータを活用した政策検討を学ぶことができます。
- **RESASではAPIを公開**しており、RESAS保有するオープンデータを利用しやすくしています。
  - RESASには、講座1-5でも紹介したAPIもあります。RESAS-APIを使うと、ウェブサイトやアプリでRESASのデータを活用することができます。
- **RESAS Communityはウェブ上の掲示板**として、RESASに関するイベントや発見の情報交換ができます。
  - ユーザー同士でRESASに関する議論ができ、各地域のRESAS関連イベントを告知、発見することもできます。

RESASに関連するウェブサービス

【出所】RESAS［内閣官房（まち・ひと・しごと創生本部事務局）］https://resas.go.jp/

□ 本講座の後で、さらにRESASを学びたい方は、RESASオンライン講座を活用してください。

人口マップ：人口推移

◆人口マップでは市町村単位での人口推移・人口予測が表示できます。

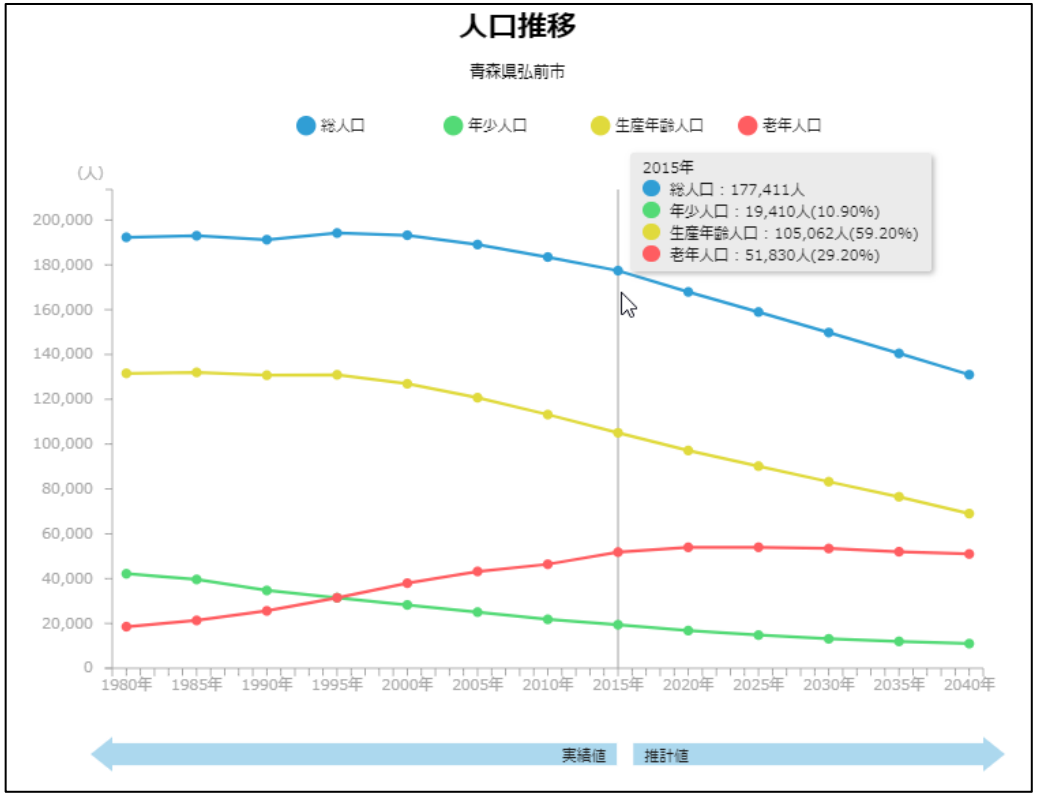
- **人口マップ**に関しては、**青森県弘前市を事例**として、人口の推移や人口の構成を確認します。
  - 弘前市は青森県の中で青森市、八戸市に次いで県内第3位の人口の市です。

【画面上】「人口マップ」＞「人口構成」を選択
【画面右上】表示レベルを指定する「市町村単位で表示する」にチェック
＞青森県・弘前市を選択＞「人口推移」ボタンをクリック

- グラフでは、青い線が総人口、緑の線が年少人口（15際未満）、黄色の線が生産年齢人口（15歳以上65歳未満人口）、赤い線が（65歳以上人口）を指しています。
- RESASではグラフにカーソルを当てると、データの数値が表示されます。
- RESASでグラフを表示した後、画面左上に表示される「×」をクリックすると、グラフを消去して画面を戻すことができます。

- 弘前市の人口は1995年にピークの19.4万人から減少しており、2015年で17.7万人となっています。さらに、RESASでは2040年における予測値13.1万人が表示されます。
  - 人口マップの人口推移では、2015年までは「国勢調査」の実績値に基づいている一方で、2020年以降は、国立社会保障・人口問題研究所が公表した推計値に基づいています。

青森県弘前市の人口推移



【出所】RESAS（人口マップ：人口構成）

人口マップ：人口ピラミッド

◆人口マップでは、市区町村単位での5歳単位の人口ピラミッドを表示できます。

- 青森県弘前市の年齢層別人口をより詳細に考察するために、**人口ピラミッド**を確認します。

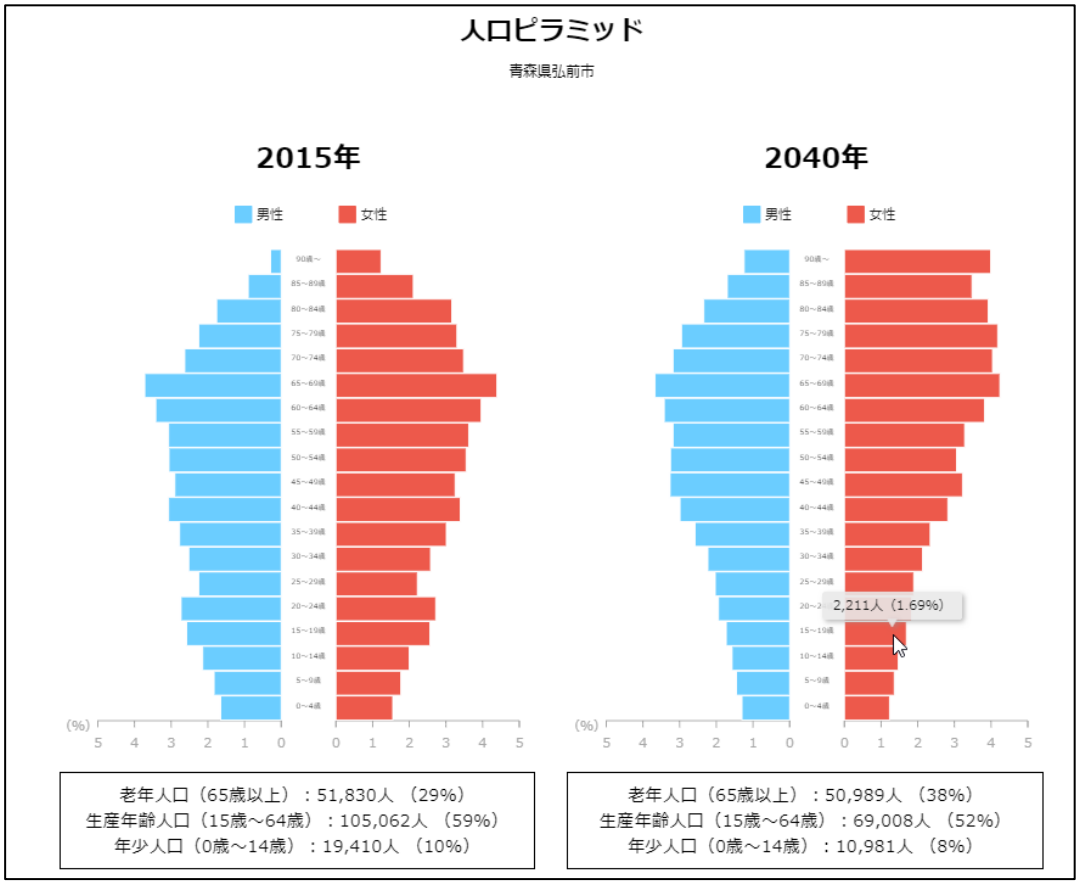
【画面上】「人口マップ」を選択＞「人口構成」を選択
【画面右上】青森県・弘前市を確認＞「人口ピラミッド」ボタンをクリック

- 人口マップでは2015年のみならず2040年の予測値に関しても、男女別5歳単位で人口ピラミッドを表すことができます。
  - 人口ピラミッドの表示対象年は、1980年から2040年まで5年単位で変更して、指定することができます。

- 各時点における「80歳以上」や「15～19歳」の人口構成を視覚的に把握することができます。
  - 人口ピラミッドにカーソルを当てると、数値データが表示されます。

□ 年齢層毎の人口推移を定量的に把握することで、教育施設・老人介護の施設の建設・運営計画を具体的に検討することができます。

青森県弘前市の人口ピラミッド



【出所】RESAS（人口マップ：人口構成）

人口マップ：自然増減と社会増減

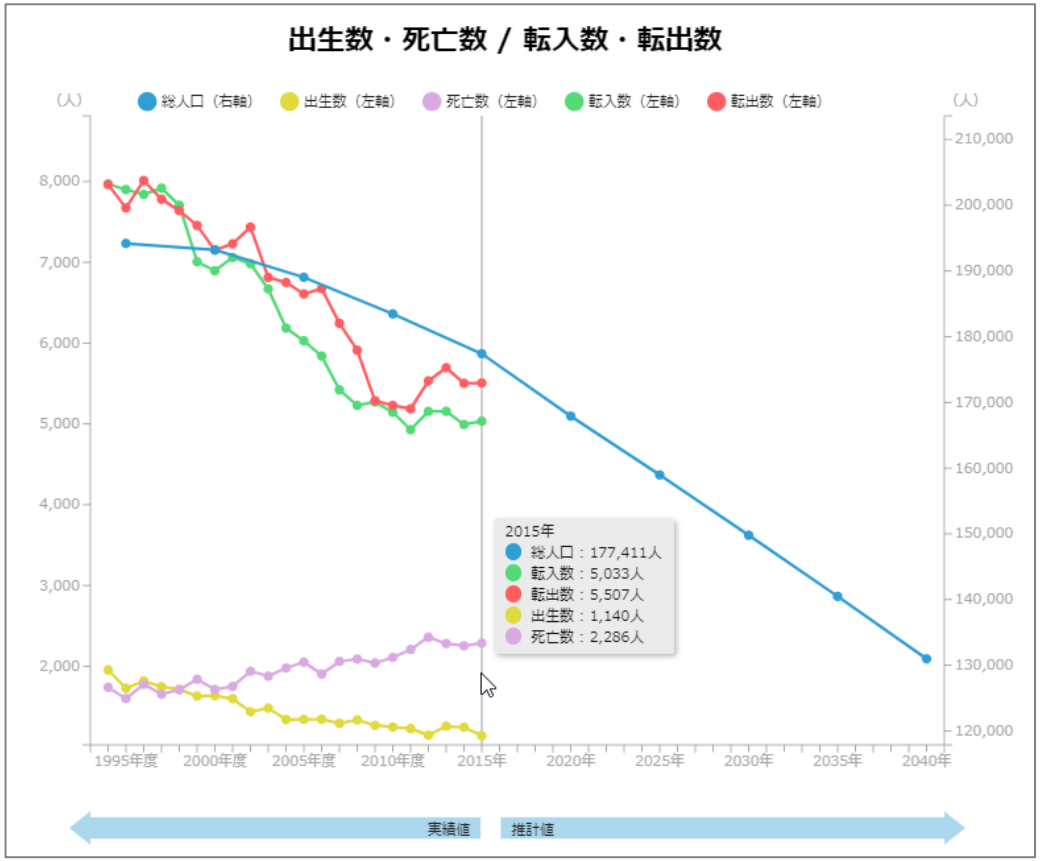
◆人口マップでは、市区町村単位での人口増減の原因を示すことができます。

- 人口マップでは人口増減を**自然増減［出生数-死亡数］**と**社会増減［転入数-転出数］**に分けることができます。
- 「自然増減」「社会増減」をさらに細分化し、1年単位で「出生数」「死亡数」「転入数」「転出数」を表示することができます。

【画面上】「人口マップ」を選択＞「人口増減」を選択
【画面右上】青森県・弘前市を確認＞
「グラフを表示」ボタンをクリック

- 自然増減は［出生数-死亡数］、社会増減は［転入数-転出数］であり、全体の人口増減は［自然増減+社会増減］で構成されます。
- 赤いバーが出生数と死亡数の差に基づく自然増減、緑のバーが転入数と転出数の差に基づく社会増減となっています。

青森県弘前市の出生数・死亡数 / 転入数・転出数



【出所】RESAS（人口マップ：人口増減）

人口マップ：年齢層別の転出超過数

◆人口マップでは年齢層別に転出先自治体、転入元自治体を把握することができます。

- 人口マップでは、**市町村単位・年齢層別に転出先自治体および転入元自治体**を示すことができます。

【画面上】「人口マップ」を選択＞「人口の社会増減」を選択
【画面右上】青森県・弘前市を確認＞「「From-to(定住人口)」」ボタンをクリック
【画面右側】「表示する年代を指定する」から
［20歳代］［30歳代］を選択

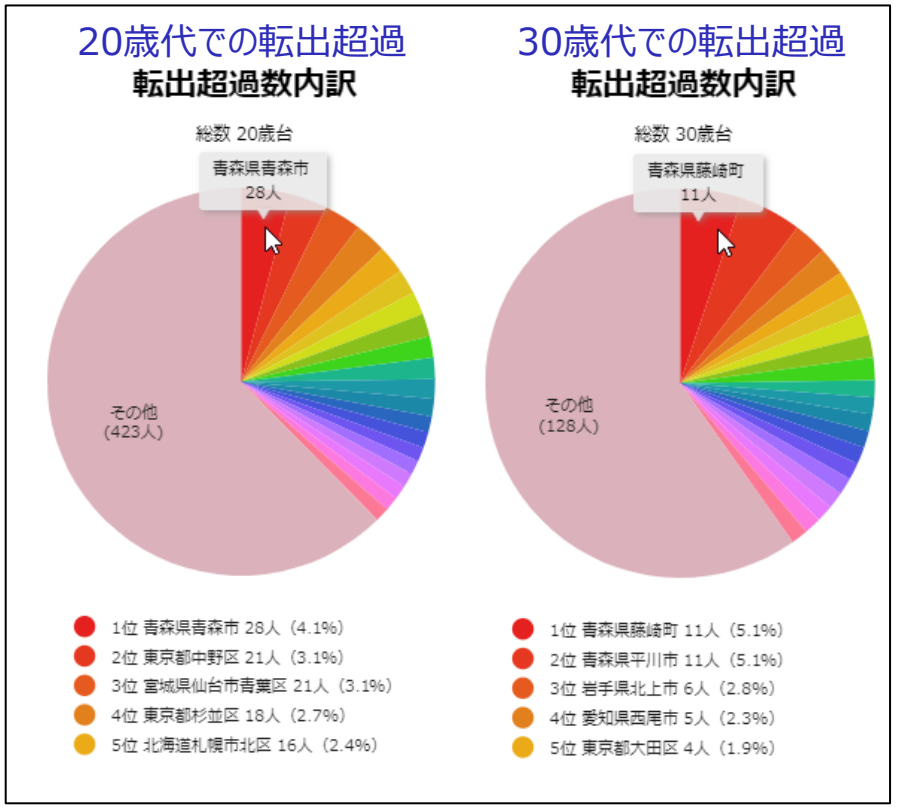
- 人口マップにおいては、転出超過（転出数-転入数）が多い自治体を市区町村単位で表示できます。

- 20歳代における弘前市からの転出超過は、青森市が最も多くなっています。次いで転出超過先として東京や仙台市が多く、弘前市より人口規模が大きい自治体への転出超過となっていることが示されています。

- 30歳代における青森県弘前市からの転出超過は、藤崎町、平川市と人口規模がより少ない自治体へ転出する傾向にあります。

□ 地域における年齢層別の転出先の傾向が分かれば、年齢層に応じた人口流出抑制政策をとることができます。

青森県弘前市の年齢層別転出超過



【出所】RESAS（人口マップ：人口の社会増減）

産業構造マップ：全産業の構造

◆産業構造マップの「全産業の構造」では、各自治体の産業別従業者数が把握できます。

- **産業構造マップ**に関しては、**鹿児島県を事例**として、産業別従業者の構成を確認します。

【画面上】「産業構造マップ」＞「全産業」＞「全産業の構造」
【画面右上】「都道府県単位で表示する」を選択、鹿児島県を選択＞「従業者数(事業所単位)」を選択＞「中分類」を指定

- 産業構造マップの「全産業の構造」の出所となっている経済センサスには、本社位置を基準とする「企業単位」と支店や工場等の事業所を基準とする「事業所単位」があります。実際に働いている場所を基準とする場合は「事業所単位」を選択します。

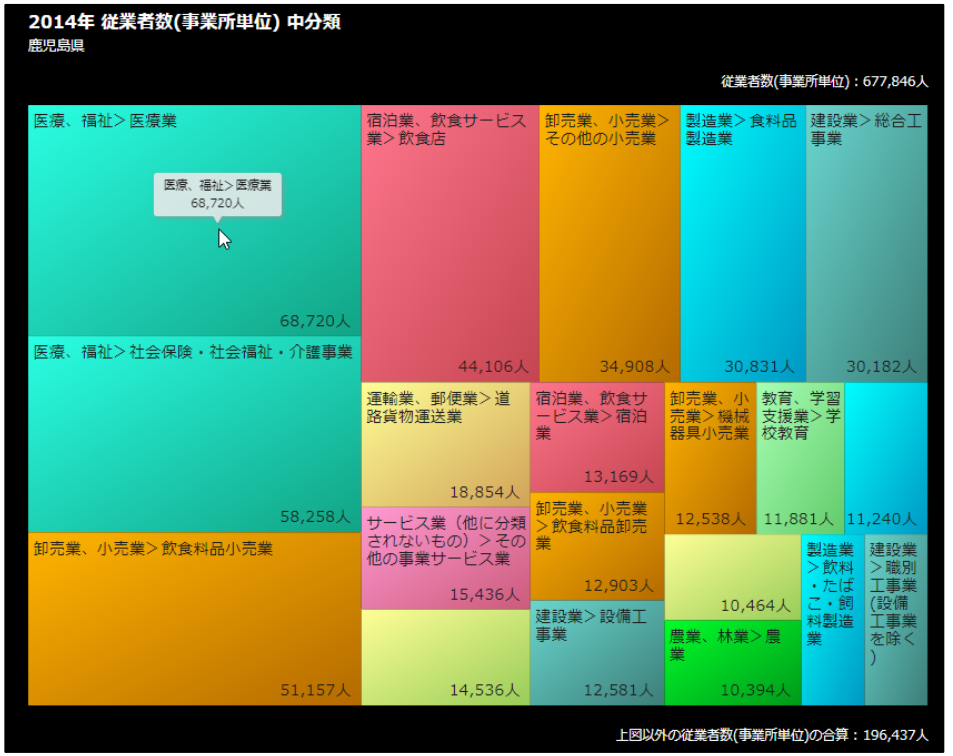
- 2014年における鹿児島県における産業別従業者数では「医療業」が最も多くなっています。

  - 医療業に次いで「社会保険・社会福祉・介護事業」「飲食料品小売業」「飲食店」の従業者数が多くなっています。

- スーパーマーケットやコンビニエンスストアの店員が含まれる「飲食料品小売業」や飲食店の店員が含まれる「飲食店」といった一般に接する従業者よりも、「医療業」「社会保険・社会福祉・介護事業」の方が従業員数が多いことが、産業構造マップから分かります。

- 一般に「医療業」「社会保険・社会福祉・介護事業」の従業者と接する機会は少ないですが、RESASによって感覚と実態のズレを確認できます。

2014年の鹿児島県における従業者数



【出所】RESAS（産業構造マップ：全産業の構造）

産業構造マップ：稼ぐ力の分析

◆産業構造マップの「稼ぐ力の分析」では、各自治体の産業別特化係数を把握できます。

- 産業構造マップの**「稼ぐ力の分析」**では、全国平均と各地域・各産業への労働投入および付加価値額の比率を**「特化係数」**として示します。

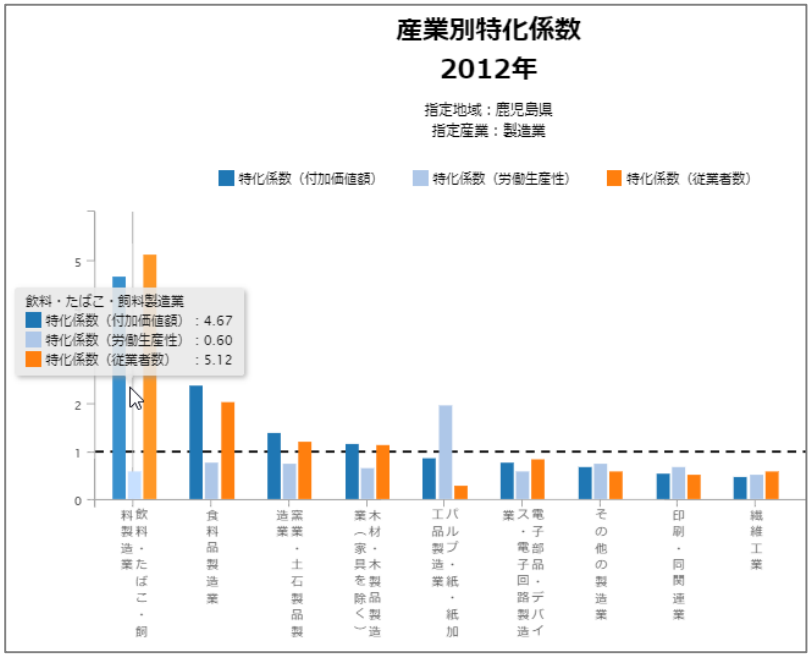
【画面上】「産業構造マップ」＞「全産業」＞「稼ぐ力分析」　【画面右上】鹿児島県の選択を確認
【画面右】「特化係数(付加価値額)【企業単位】」を選択し「製造業」「「全ての中分類」を確認、
「グラフ分析」を選択、【画面右】「特化係数(付加価値額順)で見る」を選択し、
グラフ下の水色のスクロールバーを一番左まで動かす。

- 付加価値額は、売上高から原材料費などの費用を除いた利益に相当し、日本国内の企業を横断して積み上げるとGDP(国内総生産)になります。

- 産業構造マップの「稼ぐ力の分析」は、各産業に3本の集合グラフで全国平均との比率を示しています。
  - 一番左の青色のグラフが産業のアウトプットの「付加価値額」、一番右のオレンジ色のグラフが労働インプットの「従業者数」に対応しています。

- 鹿児島県の製造業においては、「飲料・たばこ・飼料製造業」「食料品製造業」の順で、特化係数（付加価値額）が高くなっています。
  - 特化係数（付加価値額）は、全国平均に対する各自治体の相対的な付加価値額の比率を表しています。

2012年の鹿児島県における産業別特化係数



【出所】RESAS（産業構造マップ：稼ぐ力分析）

□ RESASでは詳細が分からない場合は、外部のウェブサイトやデータを活用して、より詳細なデータを収集します。

産業構造マップ：農業の構造

◆産業構造マップの「農業の構造」では、自治体の農産物販売額の構成が把握できます。

- 産業構造マップの**「農業の構造」**では、農林業センサスのデータに基づいて各自治体における農産物販売額の構成を示します。
  - 出所となっている「都道府県別農業産出額及び生産農業所得」（農林水産省）における農業は、畜産業を含んでいます。

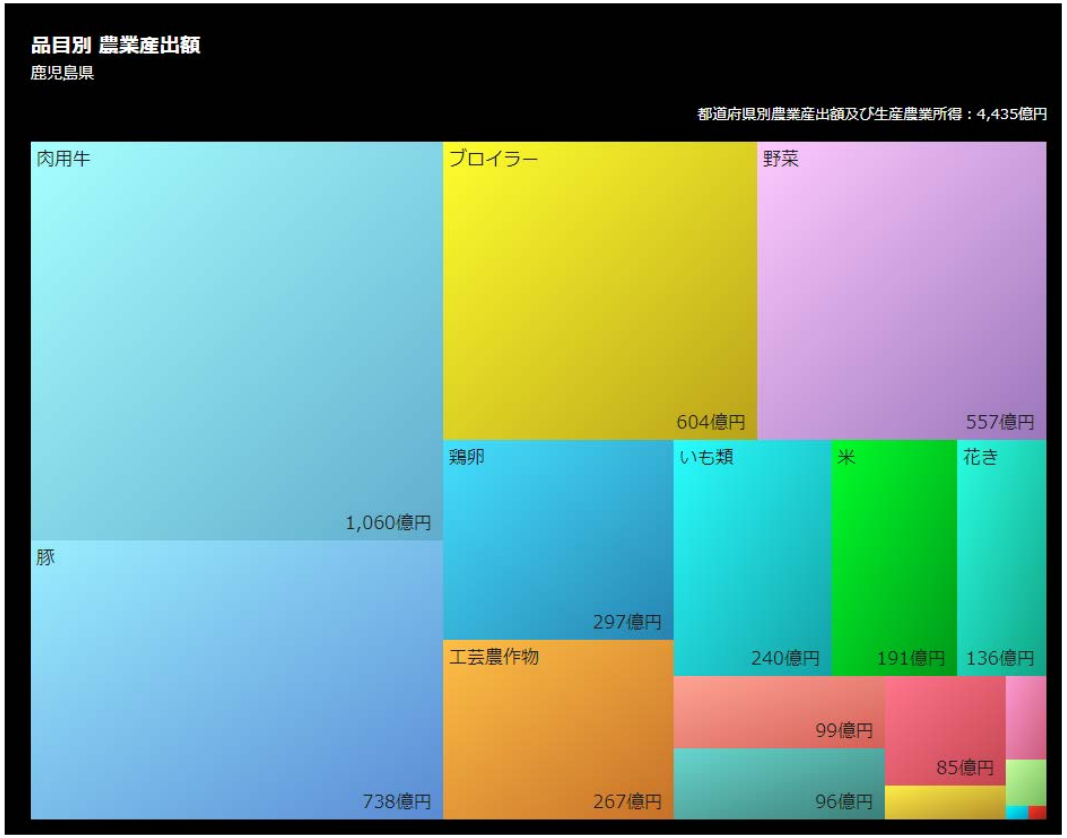
【画面上】「産業構造マップ」＞「農業」＞「農業の構造」
【画面右上】鹿児島県の選択を確認

- 鹿児島県の農業産出額は「肉用牛」「豚」「ブロイラー」「鶏卵」の順に多くなっています。
  - 鹿児島県における農業の強みは、牛、豚、鶏といった畜産業にあることが分かります。

- 2015年の鹿児島県における「肉用牛」「豚」「ブロイラー」「鶏卵」の産出額の合計は2,699億円であり、鹿児島県の農業産出額の過半を占めていることが分かります。
  - 2015年の鹿児島県における農業の産出額合計は、右図の右上に4,435億円であり、上位4種の畜産で61%を占めていることを示せます。

□ RESASによって、各地域の農業の強みを数値として確認できます。

鹿児島県における農業産出額（2015年）



【出所】RESAS（産業構造マップ：農業の構造）

産業構造マップ：消費の動向（POSデータ）

◆産業構造マップの「From-to分析（POSデータ）」では、地産地消の状況を把握できます。

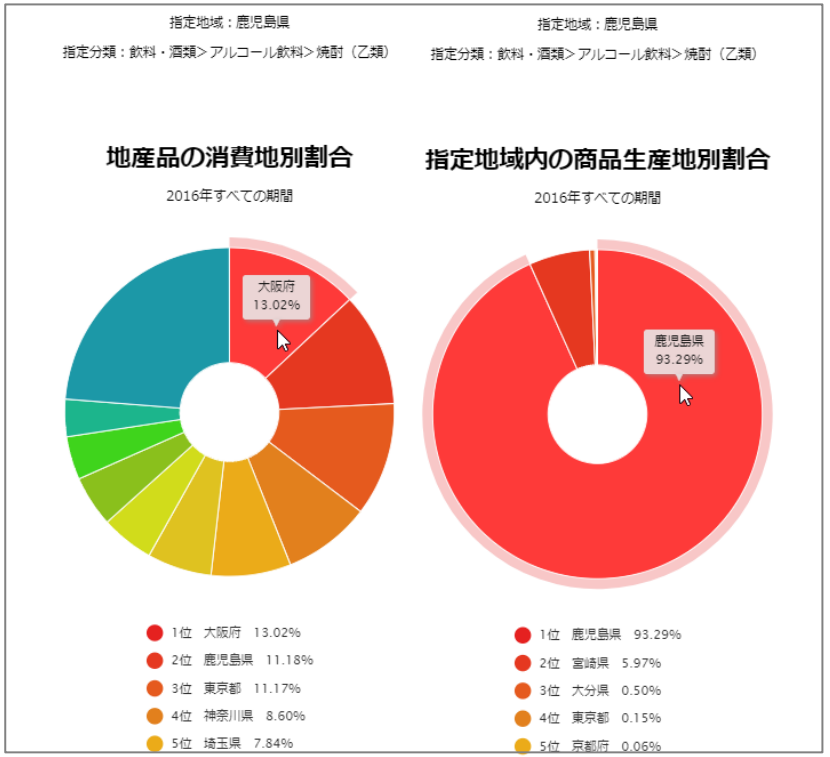
- 産業構造マップの**「From-to分析（POSデータ）」**では、スーパーやコンビニエンスストアのレジのPOSデータから、生産地と消費地の関係を都道府県別に把握できます。

【画面上】「産業構造マップ」＞「小売・卸売業（消費）」＞「From-to分析（POSデータ）」、
【画面右上】鹿児島県の選択を確認、
【画面右】表示する分類を指定する「飲料・酒類」「アルコール飲料」「焼酎（乙類）」
「地産品の消費地別割合」「指定地域内の商品生産地別割合」の
それぞれを選択して「グラフを表示」をクリック

- 「地産品の消費地別割合」によると、鹿児島県産の焼酎（乙類）は、大阪府、東京都、神奈川県、埼玉県と様々な地域で売れています。
  - 「地産品の消費地別割合」は、鹿児島県産の焼酎（乙類）が、どの都道府県で売れているかを示しています。
- 「指定地域内の商品生産地別割合」では、鹿児島県民が飲む焼酎（乙類）は93%が鹿児島県産であることが分かります。
  - 「指定地域内の商品生産地別割合」は、鹿児島県内で売れている焼酎（乙類）の製造された都道府県を示しています。

鹿児島県産の焼酎は他の都道府県でも売れていることに加えて、鹿児島県民は自県産の焼酎を地産地消しています。

鹿児島県の焼酎の消費地・地産地消の状況



【出所】RESAS（産業構造マップ：From-to分析（POSデータ））

観光マップ：目的地分析

◆観光マップの「目的地分析」から、各地域における人気のある観光地が分かります。

- **観光マップ・まち作りマップ**の利用事例として、**三重県の伊勢市および鳥羽市**を例にとって、RESASで観光振興に関する分析を行います。

【画面上】「観光マップ」＞「国内」＞「目的地分析」
【画面右上】三重県 伊勢市を選択
【画面右】「目的地検索ランキングを表示」をクリック

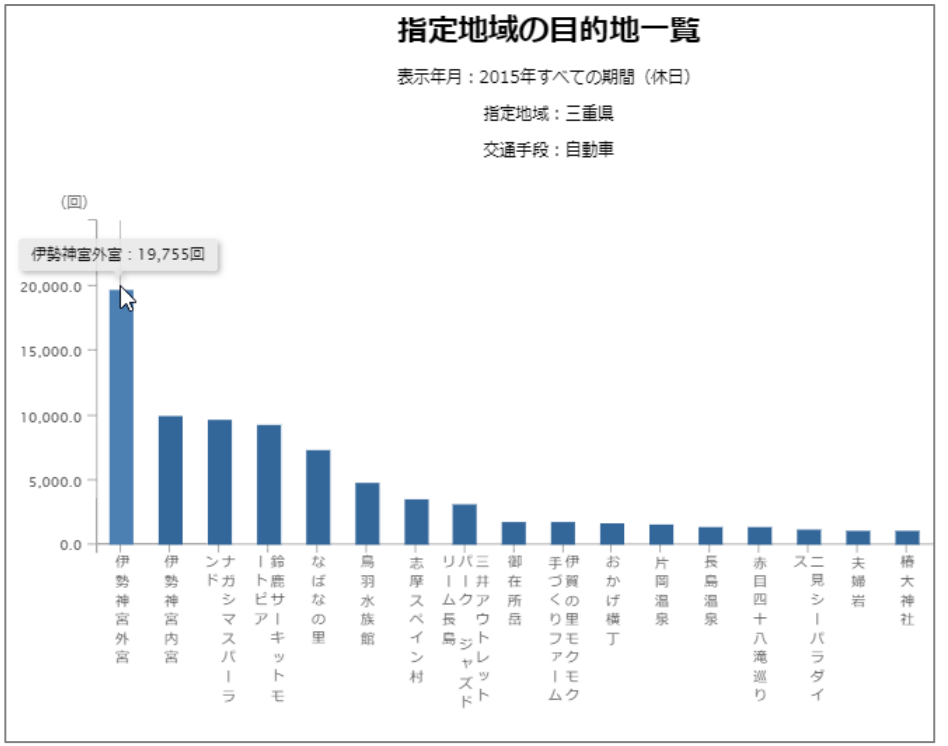
- 観光マップの**「目的地分析」**では各自治体における観光地等が、目的地としての検索頻度の高い順に表示されます。

- 「都道府県単位で表示する」をクリックして、三重県全体のランキングを表示します。
  - 目的地検索では、いったん市町村別のグラフを表示した後に「都道府県単位の表示」に変更できます。

- 伊勢神宮は、伊勢市駅から徒歩圏内にある伊勢神宮外宮（げくう）、観光通りの「おかげ横丁」からつながる伊勢神宮内宮（ないくう）の二種類があります。

- 三重県での検索ランキングでは、1位が「伊勢神宮外宮」、2位が「伊勢神宮内宮」となり、二種類の伊勢神宮が三重県で人気の高い観光地であることが分かります。

三重県における目的地一覧（目的地分析の上位）



【出所】RESAS（観光マップ：目的地分析）

まちづくりマップ：滞在人口率

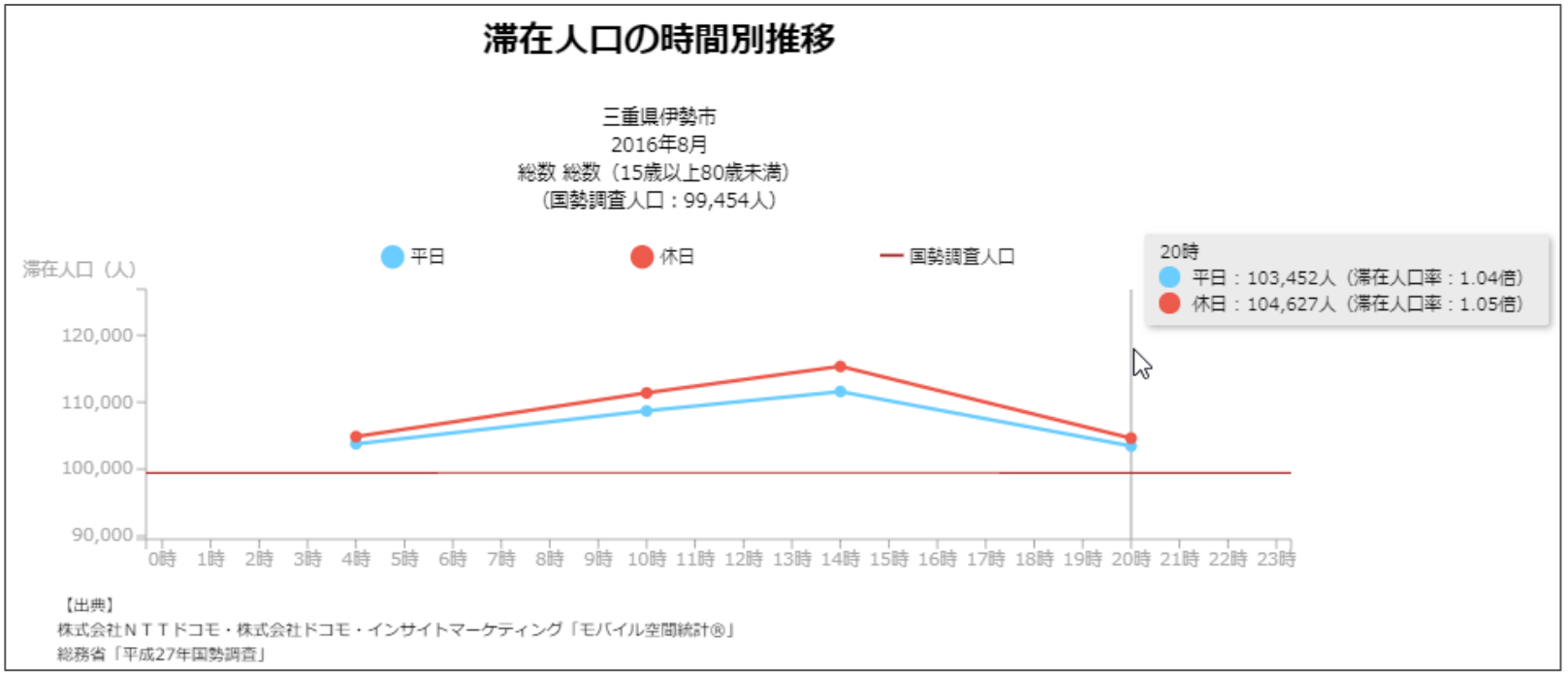
◆まちづくりマップの「滞在人口率」から、各地域の時間単位での人の移動が把握できます。

- まちづくりマップの**「滞在人口率」**を利用して、伊勢市における人の移動状況を確認します。

【画面上】「まちづくりマップ」>「滞在人口率」　【画面右上】三重県 伊勢市を選択　【画面右】「グラフを表示」をクリック

- 時間帯別の滞在人口率を見ると、2016年8月の休日に伊勢市に滞在していた人口は14時で居住人口の1.16倍に対して20時で1.04倍となっています。

三重県伊勢市における滞在人口の時間別推移



【出所】RESAS（まちづくりマップ：滞在人口率）

□ 伊勢神宮の参拝を含む旅行は、日帰りで行われているケースが多いことを示唆しています。

まちづくりマップ：流動人口メッシュ

◆まちづくりマップの「流動人口メッシュ」では、詳細な地域別に人の移動が把握できます。

- まちづくりマップの**「流動人口メッシュ」**では、より細かい範囲を区切って、時間帯別の人の集積を確認できます。
  - 「滞在人口率」は年齢層や性別を限定して人の集積を市区町村単位で把握できる一方で、「流動人口メッシュ」はより細かい地域を範囲で指定して、人の移動・集積を把握できます。

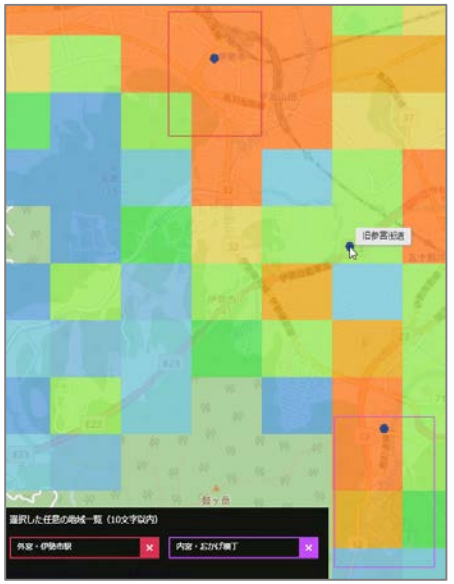
【画面上】「まちづくりマップ」＞「流動人口メッシュ」をクリック　【画面右上】三重県 伊勢市を確認、
【画面右】「メッシュなし」を「透過率80%（濃いメッシュ）」に切り替え、「メッシュを読み込む」をクリック

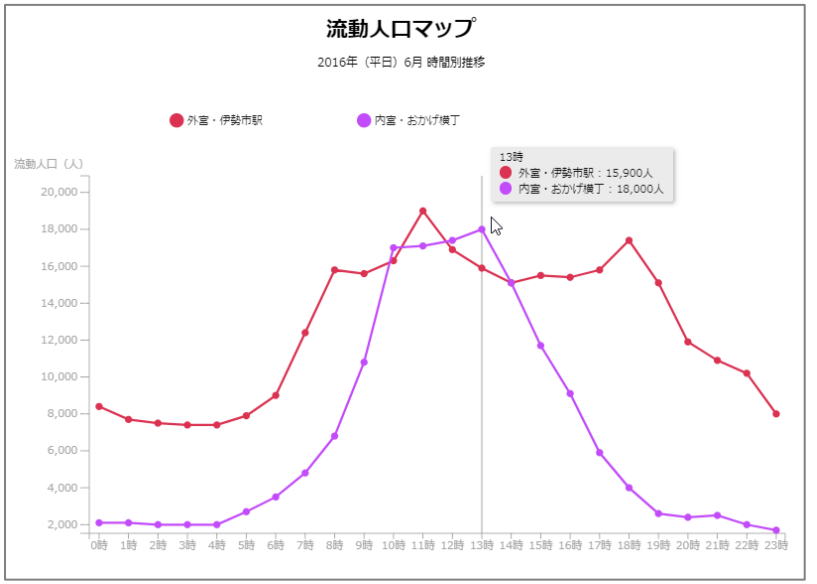
【地図上に四角形で範囲指定】地図上の「外宮・伊勢市駅」および「内宮・おかげ横丁」の該当地域を範囲指定し「Enterキー」を押し、画面左下の枠にそれぞれの範囲名を入力した後、【画面右】「グラフを表示」をクリック

- 「外宮・伊勢市駅」は伊勢市の住民や通勤者の集積が含まれることもあって、18時頃の流動は増えていますが、20時頃には減少しています。
- 時間帯別のグラフを見ると、「内宮・おかげ横丁」は、正午前後に人の集積のピークがありますが、夕方から大幅に減少していることが分かります。

- 午後14時頃からの「内宮・おかげ横丁」の流動人口の減少は、日帰り観光で中部圏、関西圏に戻るケースが多いことを示唆しています。

伊勢市近辺の流動人口メッシュと流動人口マップ





【出所】RESAS（まちづくりマップ：流動人口メッシュ）

観光マップ：From-to分析（宿泊者）

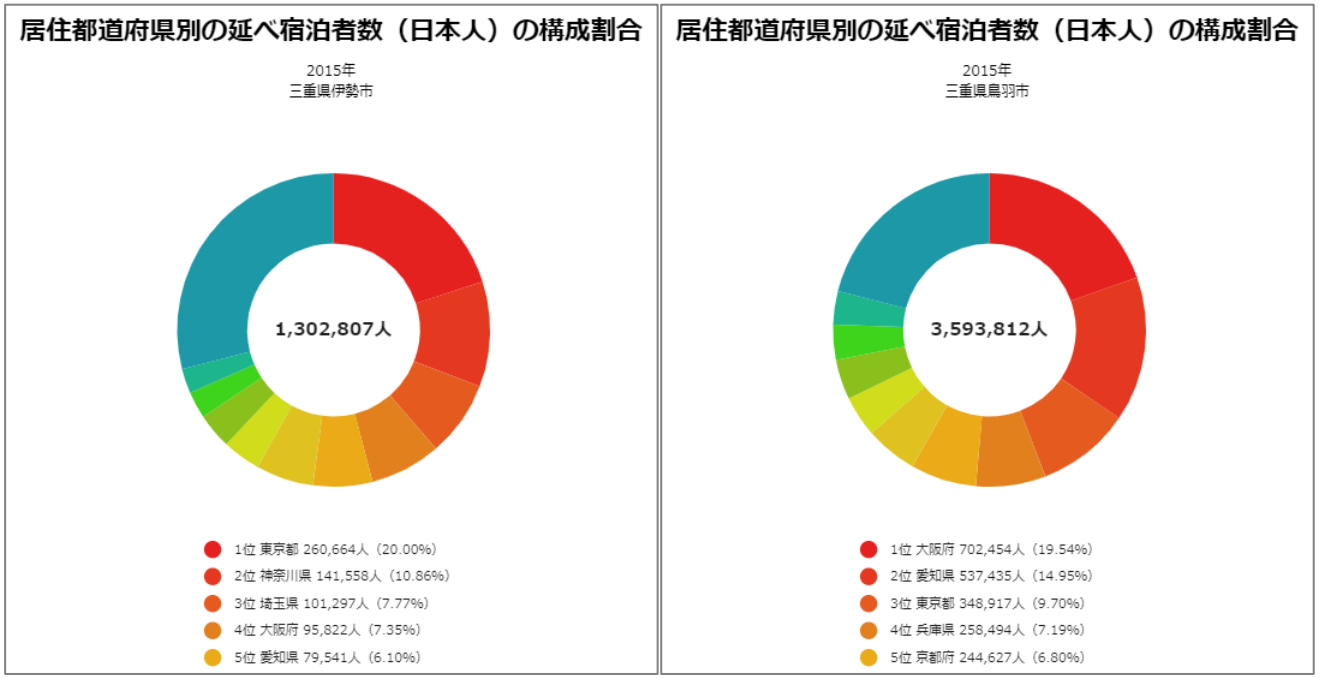
◆「From-to分析（宿泊者）」から、各市町村の宿泊者の居住都道府県が把握できます。

- 観光マップの**「From-to分析（宿泊者）」**では、各市町村の宿泊者の居住都道府県が把握できます。

【画面上】「観光マップ」＞「国内」＞「From-to分析（宿泊者）」 を選択
【画面右上】三重県 伊勢市を選択 【画面右】「居住都道府県別に見る」をクリック

- 伊勢市における2015年の宿泊者は約130万人であり、「東京都」「神奈川県」の割合が高いことが分かります。
- RESASの人口マップによると、2015年の伊勢市の人口127,817人は、鳥羽市の人口19,448人の6.6倍です。
- 一方で、伊勢市の東側に隣接する鳥羽市では2015年の宿泊者は約359万人であり、「大阪府」「愛知県」の割合が高いことが分かります。

居住都道府県別の延べ宿泊数（日本人）の構成割合［伊勢市・鳥羽市］



【出所】RESAS（観光マップ：From-to分析（宿泊者））

□ 鳥羽市の宿泊者数は多く、伊勢市と鳥羽市で宿泊者の居住都道府県の構成が大きく異なることが分かります。

RESASによる広域連携の検討

◆RESASを利用することで、広域連携での課題解決策を発見できます。

- 産業構造マップの**「稼ぐ力分析」**を用いて、伊勢市に隣接する鳥羽市の宿泊業の状況を確認します。
  - 鳥羽市は、伊勢市の東側に隣接する市であり、真珠の生産に加えて、海に近いリゾートホテル、旅館などが有名です。
  - このスライドでは鳥羽市の宿泊業の強さを確認するために、再び産業構造マップの「稼ぐ力の分析」を利用しています。

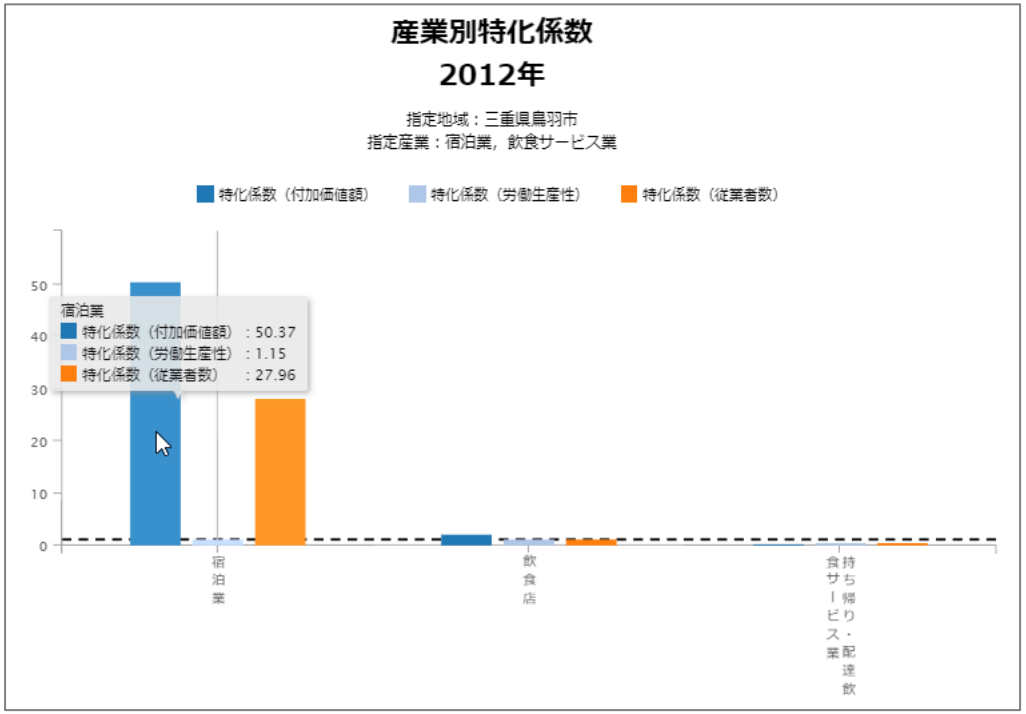
【画面上】「産業構造マップ」＞「全産業」＞「稼ぐ力分析」、
【画面右上】「市町村単位で表示する」を選択し、三重県 鳥羽市を選択
【画面右】表示産業を指定する「宿泊・飲食サービス業」を選択し、グラフ表示をクリック

- 鳥羽市の宿泊業の特化係数（付加価値額）は、50.37と非常に高い値になっています。
  - 鳥羽市における宿泊業の特化係数（付加価値額）は、全市区町村で1位となっています。

- 伊勢神宮への参拝後に、隣接する鳥羽市に宿泊するコースが確立すれば、双方の自治体にとって有益です。

□ RESASによる可視化は、地域課題の発見のみならず、広域連携の可能性を発見できます。

鳥羽市の産業別特化係数（宿泊業、飲食サービス業）



【出所】RESAS（産業構造マップ：稼ぐ力分析）

地方財政マップ：各自治体の目的別歳出

◆地方財政マップでは、各自治体の目的別支出を示すことができます。

- **地方財政マップ**に関しては、**東京都渋谷区を事例**として、自治体財政の状況を確認します。
  - RESASは地方創生を主な目的に構築されたWebアプリケーションですが、都市部の自治体にとっても有益です。
  - 地方財政マップをはじめとして、政令指定都市は区単位での詳細な情報を確認できるマップもあります。

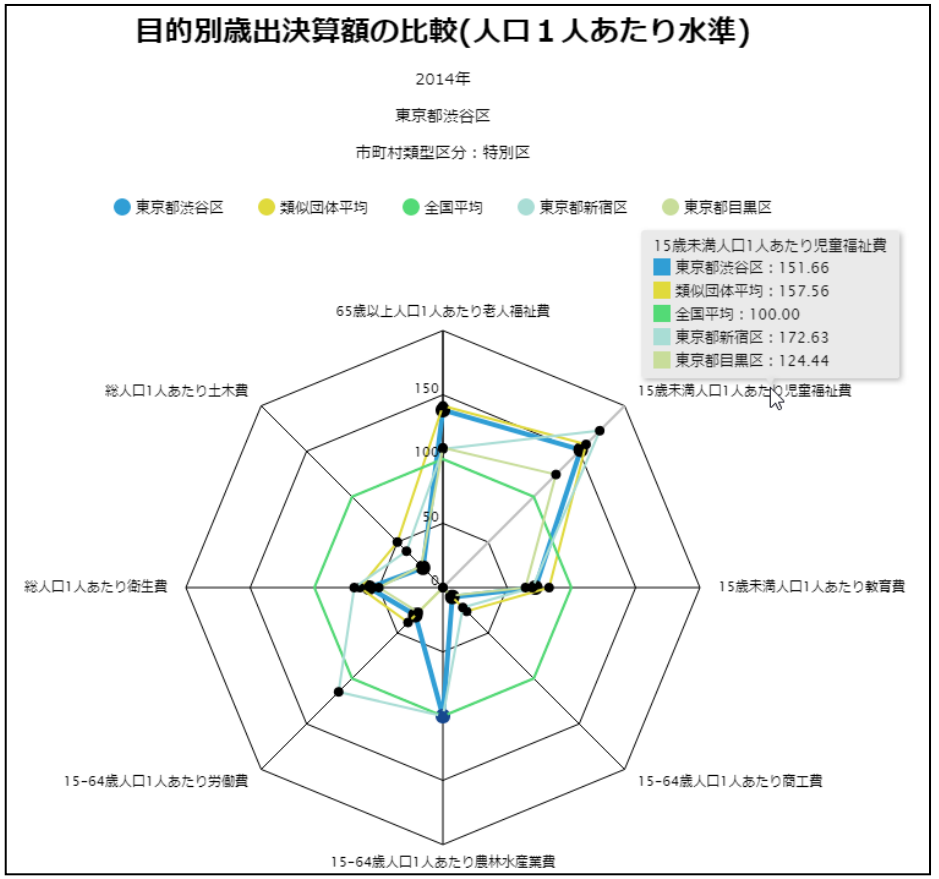
- 地方財政マップの「自治体財政状況の比較」の「目的別歳出を見る」から、各自治体の人口１人あたり目的別歳出水準を比較できます。

【画面上】「地方財政マップ」＞「自治体財政状況の比較」
【画面右上】「市区町村単位で表示する」を選択し、
東京都渋谷区を選択し、「目的別歳出を見る」を選択
【画面右】「比較地域を追加する」から新宿区、目黒区を追加

- RESASに表示されたレーダーチャートの各項目にカーソルを合わせると、全国平均を100とした相対値を見ることができます。
  - 自治体の規模や属性によって支出の傾向が異なるため、「類似団体平均」として各自治体が属するグループの平均値が表示されます。渋谷区に関する類似団体は東京23区の平均値となります。

- 児童福祉費の値に着目すると、渋谷区では152、新宿区では173、目黒区では125となっており、隣接している自治体でも支出の傾向が異なることが示されています。

目的別歳出決算額（人口1人あたり）の比較



【出所】RESAS（地方財政マップ：自治体財政状況の比較）

地方財政マップ：各自治体の財政状況

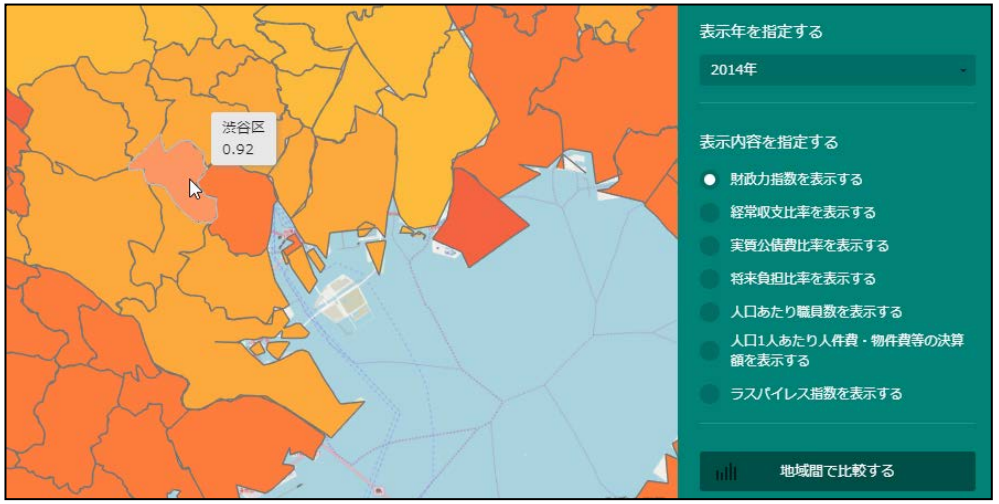
◆地方財政マップは、各自治体の財政状況を示すこともできます。

- 地方財政マップでは、各自治体の**財政状況を地図上やレーダーチャートに示す**ことができます。

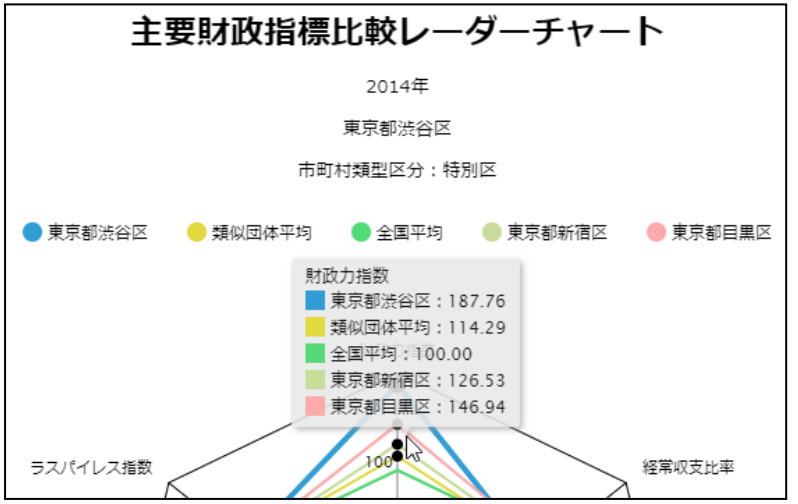
【画面上】「地方財政マップ」＞「自治体財政状況の比較」
【画面右上】「市区町村単位で表示する」を選択し、東京都 渋谷区を選択（右側の「財政力指数の表示」を確認）
【画面右】地図上で渋谷の位置にカーソルを合わせる

- 渋谷区の財政力指数は0.92となっており、必要な支出に対して自治体の収入が不足していることを示しています。
  - 財政力指数は、地方自治体における基準財政収入額を基準財政需要額で除して得た数値であり、自治体の収入に占める必要な支出の割合です。
  - RESASに表示される財政力指数は、画面上の表示年から過去3年間の平均値となっています。
  - 財政力指数が1を切る自治体の方が多数となっており、必要な支出への不足額は一般に中央政府からの交付金によって工面されます。
- 画面右下の「地域間で比較する」をクリックすると、財政力指数を相対値で他の自治体と比較することもできます。
  - 全国平均を100とした渋谷区の財政力指数は188.76で、相対的には高い財政力指数となっていることが分かります。

地図上の財政力指数の表示（実数）



レーダーチャートでの財政力指数の表示（相対値）



【出所】RESAS（地方財政マップ：自治体財政状況の比較）

RESASを出発点とする追加分析

◆RESASを出発点として、追加分析や政策・取り組みを検討することも可能です。

● RESASに**「データをダウンロード」**のボタンが表示されるものでは、データをダウンロードして追加分析を行えます。

- 前スライドの「地域間で比較する」のレーダーチャートの画面右側に「データをダウンロード」のボタンがあります。

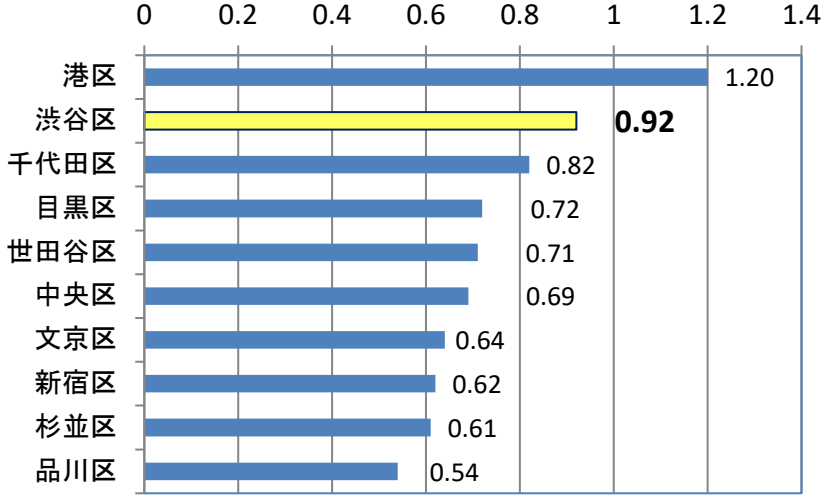
● ダウンロードデータは、RESASで閲覧していた情報の一括データがCSVに格納されており、Excelなどで開いて**自由なデータ分析**を行うことができます。

- データをダウンロードすれば、利用者一人一人が作成したい独自の図表作成や外部データと組み合わせた分析ができます。

● 着目している**指標が高い自治体を検索・調査**することで、自地域における政策・取り組みの参考にすることができます。

- RESASから直接、具体的な政策・取り組みを発見することは困難ですが、着目している指標に関して業績の良い自治体の環境や政策を調査することで、自地域の政策・取り組みの参考にすることができます。

東京23区の財政力指数（2014年：上位10区）



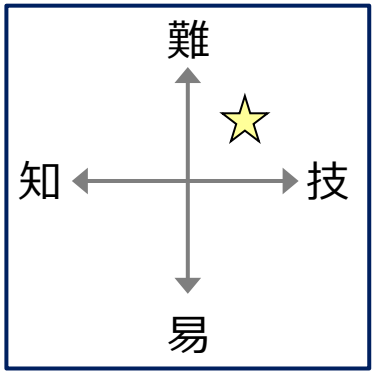
財政力指数（2014年）全国上位5位

順位	都道府県	市区町村	財政力指数
1	愛知県	飛島村	2.07
2	北海道	泊村	1.88
3	山梨県	山中湖村	1.81
4	青森県	六ヶ所村	1.64
5	長野県	軽井沢町	1.49

【出所】愛知県飛島村 http://www.vill.tobishima.aichi.jp/

◘ RESASは、本格的な分析や政策・取り組みの調査・検討の出発点としても利用することができます。

総務省　ICTスキル総合習得教材



［コース4］オープンデータ・ビッグデータ利活用事例

4-3：プログラミングによるビッグデータの分析（R）

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_4_3.pdf

	1	2	3	4	5
［コース1］データ収集					
［コース2］データ蓄積					
［コース3］データ分析					
［コース4］データ利活用			●		

本講座の学習内容［4-3：プログラミングによるビッグデータの分析（R）］

【講座概要】

- プログラミング言語のR、統合環境のRStudioの概要と画面構成と基本操作を説明します。
- RStudioを用いたExcelファイルの読み込み方法、Rのパッケージ利用方法を示します。
- Rはパッケージを利用することで、様々な分析や出力が簡単に行えることを示します。
- 一般電気事業者が公開している1時間単位の消費電力ビッグデータをRを用いて分析します。
- 【参考2】においては、Rによるe-Stat APIの利用手順を示します。

【講座構成】

実習

- [1] RとRStudioの概要
- [2] Rにおける回帰分析・パッケージ利用
- [3] Rによるビッグデータの分析
- 【参考1】RとRStudioのダウンロード・インストール
- 【参考2】Rによるe-Stat APIの利用

【学習のゴール】

- ✓ 統計分析に適したプログラミング言語のRと統合環境のRStudioの概要を把握する。
- ✓ RStudioにおける画面構成、基本操作、パッケージの利用方法を把握し、Rにおける基礎的なプログラミングができるようになる。
- ✓ Rを用いたビッグデータ分析を通して、定量的な効果測定、予測、発見の事例を理解する。

統計分析ソフトウェアRとRStudio

◆この講座では、統計分析用ソフトウェア（プログラミング言語）のRの実習を行います。

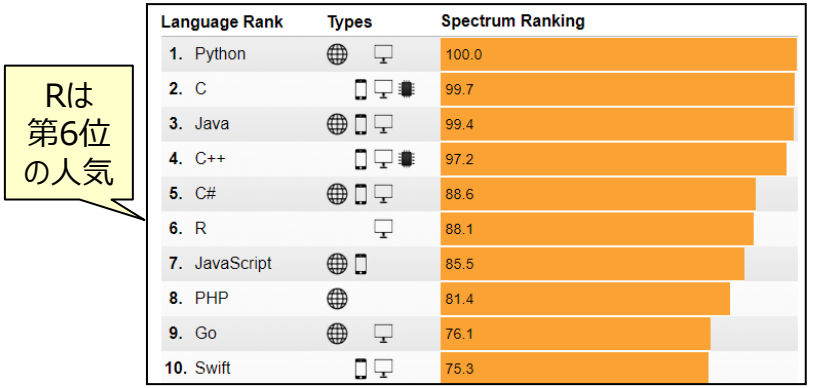
- この講座では、様々な分析に利用されている**R**（アール）を説明し、その活用を実習形式で行います。
  - Rは、Windows、Macintosh、Linuxにインストールできる無料のソフトウェアであるとともにプログラミング言語です。
- **Rは、データ分析に特化したプログラミング言語**で、データ分析の初心者から専門家まで幅広く人気があります。
  - 様々なソフトウェアの制作に利用されるC言語やJavaといった汎用プログラミング言語と異なり、Rはデータ分析がしやすい設計になっています。
  - 米国電気電子学会が人気のあるプログラミング言語を示した「The Top Programming Languages 2017」において、Rは第6位になっています。
- **RStudio**は、Rを快適に利用することができる**統合開発環境**です。
  - 統合開発環境（IDE: Integrated Development Environment）は、一つのソフトウェアの中に入力欄、出力欄、データ欄等が統合されて表示されることで、プログラミング等による開発を行いやすくする環境です。
  - RStudioは、無料で利用できるオープンソース版と優先的なサポートが受けられる商用ライセンスがあります。

統計分析ソフトウェアRのロゴ

© 2016 The R Foundation.

【出所】The R Foundation
https://www.r-project.org/

The Top Programming Languages 2017の順位



【出所】米国電気電子学会（IEEE）
https://spectrum.ieee.org/static/interactive-the-top-programming-languages-2017

統合開発環境RStudioのロゴ

RStudio is trademarks of RStudio, Inc

【出所】RStudio, Inc
https://www.rstudio.com/

□ RおよびRStudioのダウンロード・インストール方法は、本教材の【参考1】に記載しています。

■ Rのプログラミングコードを記入したファイルおよび該当する入力コード番号は、各スライドの右上に示しています。

http://www.soumu.go.jp/ict_skill/dc/ict_4_3data_code.zip

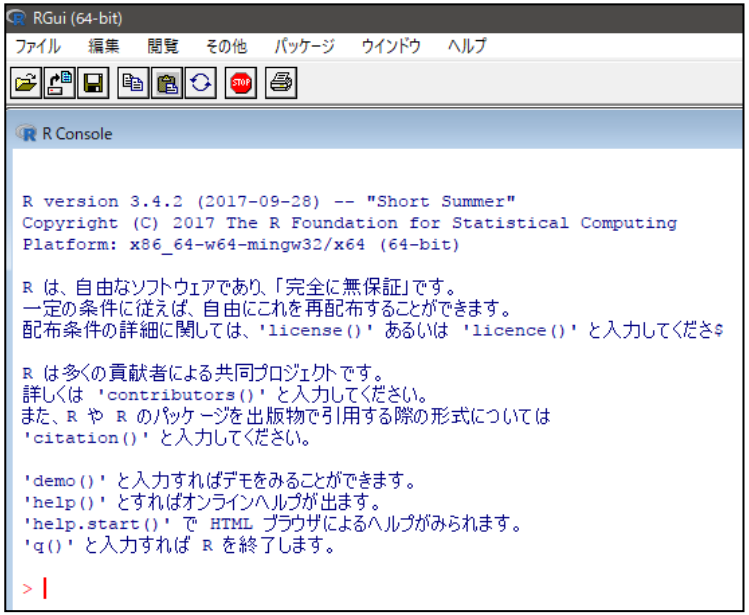
Rの起動と基本操作

◆まず、RStudioを使わずにRを直接操作して、プログラミングと出力の関係を確認します。

- RおよびRStudioのインストール後は、右下のようなショートカットアイコンが表示されます。
  - 「R i386」は32ビット版のRを指し、「R x64」は64ビット版のRを指します。Windowsの場合は、利用しているWindowsが32ビット版なら「R i386」、64ビット版なら「R x64」を使ってください。利用しているWindowsが、32ビットか64ビットか分からない場合は、どちらでもプログラムが動く「R i386」を使ってください。

- まず、RStudioを使わずに直接Rを操作するために、Rのショートカットアイコンをクリックして起動します。
  - Rの基本部分は日本語化済みで、初期画面にはRのライセンスに関する日本語での説明が表示されます。
- Rの直接操作、プログラミング体験として、中央下の**枠内の黒字の部分の入力し、出力を確認**します。
  - Rでは「＃（番号記号、ナンバーサイン、ハッシュ）」の右側はプログラムとして読み込まないので、「＃」の右側には日本語を含めて説明書きやコメントを記入することができます。

Rの初期画面の表示



Rでの基本演算［入力コード01］

```
#足し算としての「1+2」
1+2
#Rで変数を作る場合は
#「変数名 <- 変数の中身」で入力
#ｘに10、yに20を入力
x<- 10
y<- 20
#xとyの足し算としてのz
z=x+y
#変数名を入力すると、変数の値を出力
z
#全体を()でくくると、計算と同時に出力
(zz=x*y)
```

Rの出力

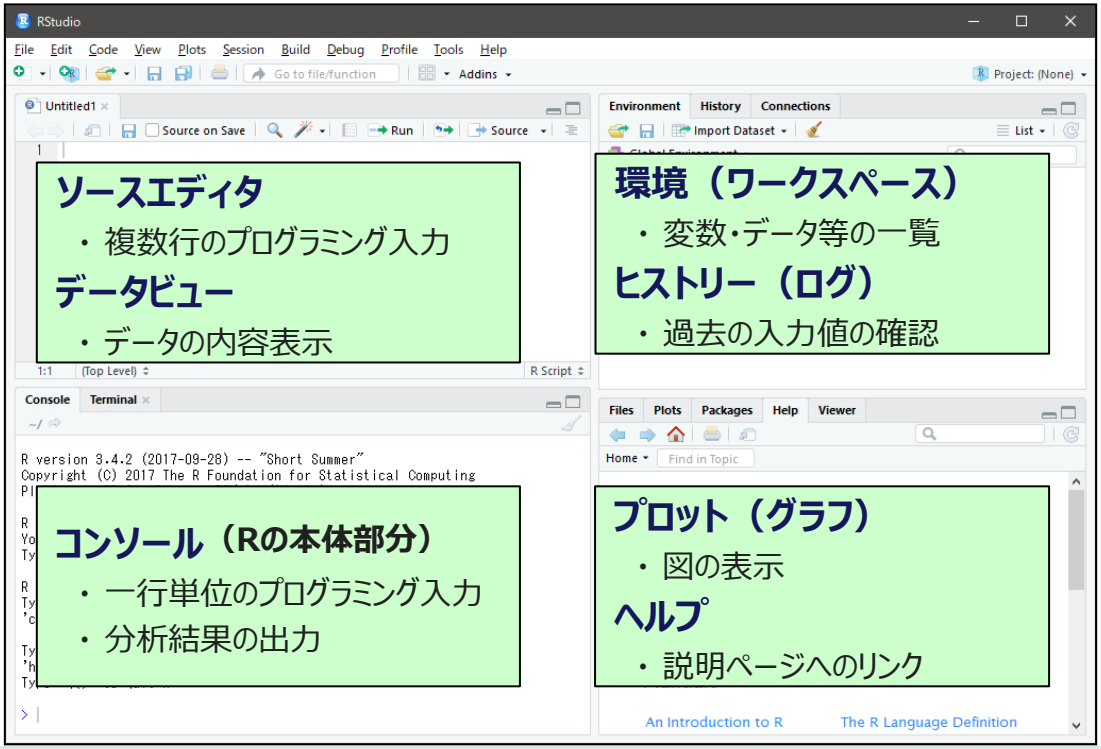
```
> #足し算としての「1+2」
> 1+2
[1] 3
> #Rで変数を作る場合は
> #「変数名 <- 変数の中身」で入力
> #xに10、yに20を入力
> x<- 10
> y<- 20
> #xとyの足し算としてのz
> z=x+y
> #変数名を入力すると、変数の値を出力
> z
[1] 30
> #全体を()でくくると、計算と同時に出力
> (zz=x*y)
[1] 200
```

RStudioの画面構成

◆RStudioは分割した画面構成によって、Rのプログラミングを効率的に行えます。

- RStudioのショートカットアイコンクリックすると、**分割された画面構成のRStudio**が起動します。
  - RStudioには、公式の日本語版はありませんが、初歩的な英単語の知識で概ね読むことができます。また、ウェブ上の無料翻訳サービスを活用すれば、英語が苦手でもRStudioの利用に支障はありません。
  - もし画面の左側が縦に分割されていない場合は、画面上側のメニューの左端にある[File]→[New File] →[R Script] を選択してください。
- RStudio内では分割された各パネルで**入力欄**、**出力欄**、**データ表示欄**、**グラフ表示欄**と機能が分けられています。
  - RStudioでは分割された各パネルにタブ（つまみボタン）が付いており、パネル内の表示内容や表示対象を変えることができます。
  - RStudioの画面構成は、メニューの[Tools]→[Global Options] →[Panel Layout]から、利用者の好みに合うようにカスタマイズできます。

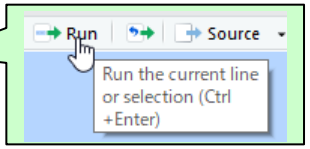
初期設定におけるRStudioの画面構成（主なタブの内容）



RStudioへの入力と画面出力

◆RStudioを使うと、変数データー覧やグラフを確認しながら、プログラミングができます。

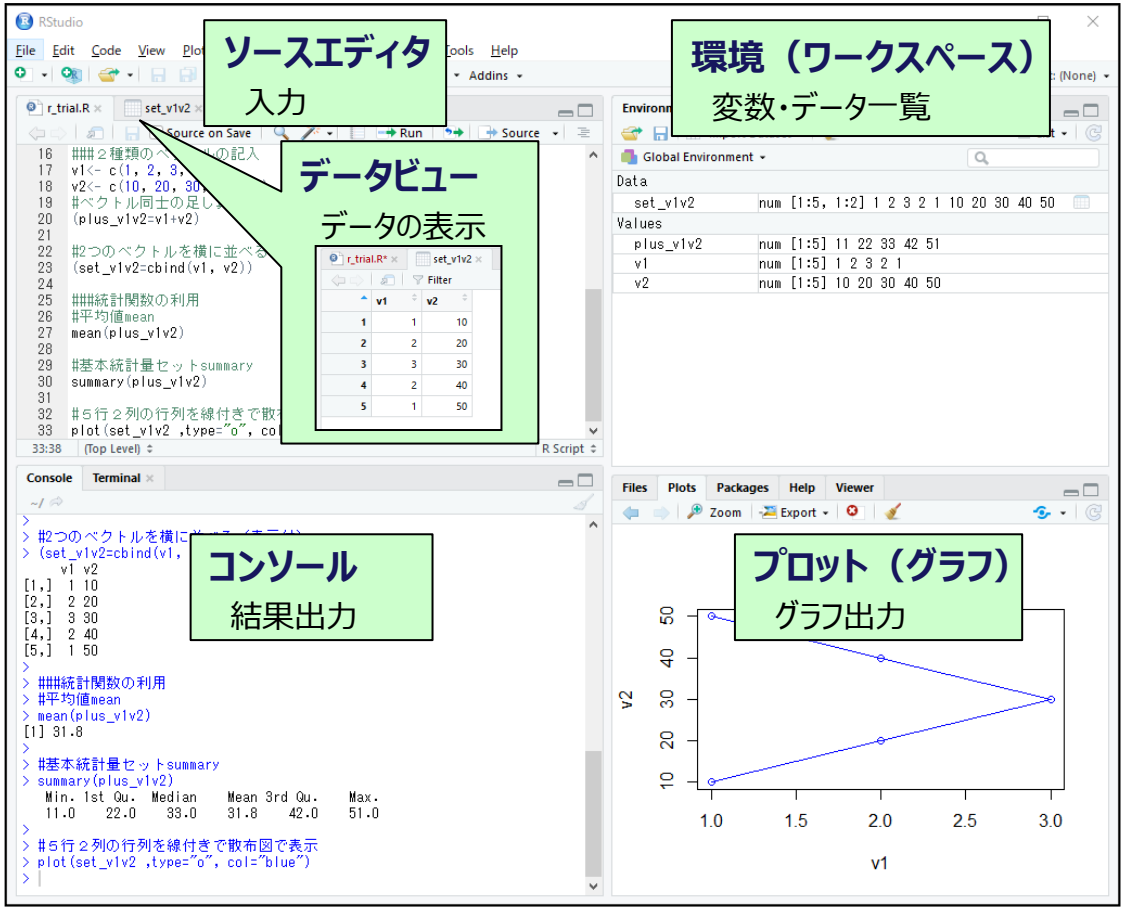
- 右下の画像では、ベクトル形式のデータを操作、線付きの散布図（グラフ）の描画を行っています。
  - RStudioでは、4分割の左上にあるソースエディタから実行したいコードの範囲を選択後、「Run」のボタンをクリックしてください。
  - ベクトルは、数値を横（行）または列（縦）に並べたものを指し、数値を束ねたもののイメージです。
  - Excel同様、括弧内に対象を指定して定められた処理をするものを関数といいます。例えば、mean関数は平均値を導出します。



基本統計量の導出等［入力コード02］

```
###２種類のベクトルの記入
v1<- c(1, 2, 3, 2, 1)
v2<- c(10, 20, 30, 40, 50)

#ベクトル同士の足し算（表示付）
(plus_v1v2=v1+v2)

#2つのベクトルを横に並べて行列作成（表示付）
(set_v1v2=cbind(v1, v2))

###統計関数の利用
#平均値mean
mean(plus_v1v2)

#基本統計量セットsummary
summary(plus_v1v2)

#「set_v1v2」を線付きで散布図で青で表示
plot(set_v1v2 ,type="o", col="blue")
```

RStudioの4分割画面の表示

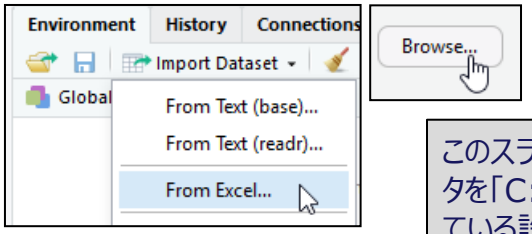


RStudioにおけるExcelファイルの読み込み

◆RStudioでは、Excelデータをマウス操作で簡単に読み込むことができます。

- RStudioの標準設定における右上のパネルの**[Import Dataset]から外部のデータを読み込み**ます。
- 実習用Excelファイルの取り込む場合は[From Excel]→[Browse]とクリックし、「tokyo_drink.xlsx」の選択後、プレビューでデータの内容を確認してから [Import]をクリックしてください。
  - Excelファイル内の分析用データは1行目に変数名、2行目以降に一行ずつ個別の標本のデータが入っている形式にしておきます。
  - ファイル名が日本語だと取り込む際にエラーになることがあるので、半角英数字のファイル名にしておくと、読み込み時のエラーの心配がありません。一方、Excelファイルの中身の各セルに入っているデータは、文字データの列であれば日本語が含まれていても問題ありません。
  - データがプレビューに表示されている状態では、［Code Preview］にデータと読み込みに対応するコードが表示されます。このコードをコピーして、ソースエディタに貼りつけることで、次回以降の同じデータ読み込みをする際にはマウスによる操作は不要でコード内で行えます。

Excelファイルの指定と[Browse]

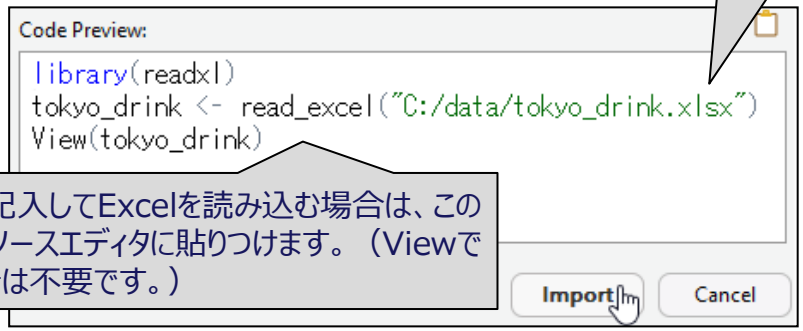


プレビューによるデータ内容の確認

このスライドでは、実習用データを「C:/data/」下に設置している設定で説明します。

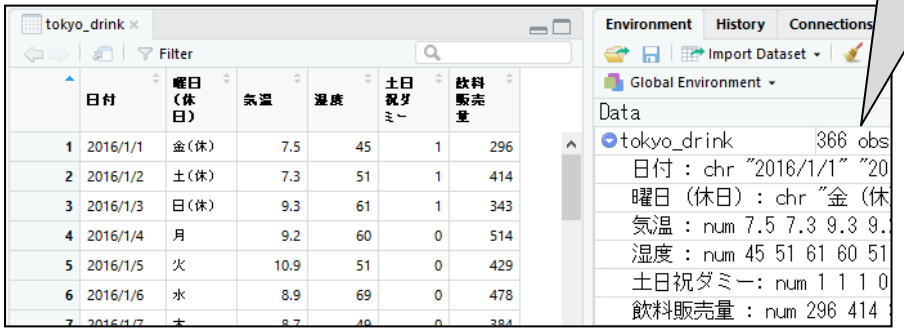
講座3-4の「Excelによる回帰分析」で利用したデータと内容は同じです。

コードとしての表示と[Import]



コードに記入してExcelを読み込む場合は、このコードをソースエディタに貼りつけます。（Viewで始まる行は不要です。）

RStudio内に取り込んだExcelデータの表示



Rにおける回帰分析

◆Rでは読み込んだデータに対して、1行のコードで回帰分析が実行できます。

- RStudioから読み込んだExcelファイルは、**データフレームと呼ばれるデータの形式**となり、データフレーム内の各列は「データフレーム名$列名（変数名）」で指定することができます。
  - RStudioの標準的なExcelファイルの読み込み設定では、Excelファイル上のデータの1行目が、Rにおける列名（変数名）となります。
- Rにおける回帰分析は、「lm(被説明変数~ 説明変数1 + 説明変数2 +...)」という1行のコードで実行できます。

プログラミングをしやすいように短いデータフレーム名「td」へとコピーするために「td<-tokyo_drink」と入力します。

回帰分析の結果を「lm_result」という名前のデータ(リスト形式)として格納するために、
「lm_result<-lm(td$飲料販売量 ~ td気温 + td湿度 + td$土日祝ダミー)」と入力します。

- 回帰分析の結果データに対して、summary関数を使うと、回帰分析の結果が表示されます。
  - summary関数は、対象にデータ自体を指定すると基本統計量を表示しますが、回帰分析の結果を指定すると、分析結果を表示します。

講座3-4と同一の回帰分析［入力コード03後半］

```
#データフレーム名を短くして、新たにデータフレームを作成
td<-tokyo_drink
#td内の飲料販売量を被説明変数、気温と湿度と土日祝ダミーを説明変数として回帰分析
lm_result<-lm(td飲料販売量~td$気温+td$湿度+td$土日ダミー)
#決定係数などの結果確認
summary(lm_result)
```

Rの統計的検定においては
両側5%で有意なら「*」、
両側1%で有意なら「**」、
両側0.1%で有意なら「***」
が表示されます。

環境（ワークスペース）における表示

Data	
data_ols	50 obs. of 4 variables
lm_result	List of 12

講座3-4におけるExcelの分析ツールでの回帰分析と同じ結果です。

Summaryによって示される回帰分析の説明変数の出力

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  374.9633    14.5686  25.738   <2e-16 ***
td$気温        5.8803     0.4914  11.965   <2e-16 ***
td$湿度        0.4903     0.2318   2.115   0.0351 *
td$土日祝ダミー -86.7386    6.9676 -12.449   <2e-16 ***
```

Rにおけるパッケージの利用

◆Rでは、ウェブ上に公開されているプログラム（パッケージ）を利用することができます。

- **公開されているRのプログラム（パッケージ）**を利用すると、簡単な入力で様々な出力・分析を行うことができます。
  - Rにおいて、有用なプログラミングコードを配布用にとりまとめて公開しているものを「パッケージ」といいます。
  - インターネット上のCRANに保存されているパッケージを初めて使う場合は、コードに「install.packages(“パッケージ名”)」と入力し、PC内にパッケージをダウンロード・インストールしてください。（一度、PCにインストールすれば、2回目以降のコードへの記載は不要です。）
  - CRANは原則として英語での表記ですが、CRAN内を検索してパッケージを探すこともできます。（https://cran.ism.ac.jp/search.html）
- PCにインストール済のパッケージは、コードに「library(パッケージ名)」と入力した後に使うことができます。
  - libraryは、「図書室」「書庫」「蔵書」を意味し、Rではlibrary関数によって、インストール済のパッケージを読み込みます。
- 複数の回帰分析の結果を並べて表示して、比較したい場合には、memiscパッケージが便利です。
  - memiscパッケージに含まれるmtable関数に対して、回帰分析の結果データをコンマで区切って指定してください。

パッケージを利用した回帰分析の比較表示［入力コード04］

```
#回帰式から気温と湿度の各変数を外して、それぞれ2変数で回帰して結果をlm_res2、lm_res3に格納
lm_res2<-lm(td$飲料販売量~td$気温+td$土日祝ダミー)
lm_res3<-lm(td$飲料販売量~td$湿度+td$土日祝ダミー)

#パッケージ「memisc」のインストールと利用宣言（installで始まる行は1度目のみ[要インターネット接続]）
#install.packages("memisc")
library(memisc)

#パッケージmemisc内のmtable関数を利用
#3つの回帰分析の結果を並べて表示
mtable(lm_result, lm_res2, lm_res3)
```

memiscパッケージによる回帰分析の比較出力

◆Rはパッケージを利用することで様々な出力、高度な分析を簡単に実行できます。

- memiscパッケージに含まれるmtable関数では、回帰分析の結果を並べて分かりやすく表示できます。

memiscパッケージ（mtable関数）による回帰分析の比較出力

回帰分析の結果表示において、(Intercept)は、切片の高さを表し、説明変数の値が全て0の場合における被説明変数の予測値に対応します。

説明変数に関する出力において、括弧のない値は「推定係数」を表し、括弧に入った値は「推定係数の標準誤差」を表しています。

「R-squared」は決定係数を意味し、講座3-4においても示したように0以上1以下の値をとる回帰分析の当てはまり度合いの指標です。

「adj. R-squared」は自由度調整済み決定係数です。

Console　Terminal ×

C:/syda/

	lm_result	lm_res2	lm_res3
(Intercept)	374.963***	400.563***	376.813***
	(14.569)	(8.151)	(17.185)
td$気温	5.880***	6.394***	
	(0.491)	(0.429)	
td$湿度	0.490*		1.860***
	(0.232)		(0.238)
td$土日祝ダミー	-86.739***	-87.746***	-84.283***
	(6.968)	(6.984)	(8.216)
R-squared	0.518	0.512	0.328
adj. R-squared	0.514	0.510	0.324
sigma	62.689	62.989	73.953
F	129.849	190.710	88.521
p	0.000	0.000	0.000
Log-likelihood	-2031.898	-2034.147	-2092.882
Deviance	1422640.504	1440228.237	1985284.010
AIC	4073.796	4076.293	4193.765
BIC	4093.310	4091.904	4209.375
N	366	366	366

決定係数を横並びで比較することで「回帰分析のあてはまり」に重要な役割を持つ説明変数を把握しやすくなります。

Rによる消費電力ビッグデータの分析

◆このパートではRを用いて、ビッグデータ（1時間単位の消費電力）の分析事例を示します。

- Rをはじめとするプログラミングでは、**大容量のビッグデータのデータ処理・分析を行うことができます。**
  - プログラミングによる分析作業は、マウス等による対象データ範囲の指定が不要であるため、一般に標本数やデータサイズに依存しません。
  - 64ビット版のRでは、メインメモリの容量を上限としてデータを格納することができ、GB（ギガバイト）単位のデータ処理が可能です。
- このパートでは、9つの電力会社管内における**1時間単位の消費電力データ・気象データを分析**に利用します。
  - 沖縄電力を除く9つの一般電気事業者が運営する電力系統は、電力を融通し合う連系線によって近隣の電力系統と結ばれており、本分析は効率的な電力の融通の検討にも利用できます。
  - 分析に利用する消費電力データおよび気象に関するデータは、誰でも下記のURLからダウンロードすることができます。

分析実習用データの提供主体、データにつながるURL

提供主体	分析用データにつながるURL
北海道電力	http://denkiyoho.hepco.co.jp/area_download.html
東北電力	http://setsuden.tohoku-epco.co.jp/download.html
東京電力	http://www.tepco.co.jp/forecast/html/download-j.html
中部電力	http://denki-yoho.chuden.jp/
北陸電力	http://www.rikuden.co.jp/denki-yoho/#download

提供主体	分析用データにつながるURL
関西電力	http://www.kepco.co.jp/energy_supply/supply/denkiyoho/
中国電力	http://www.energia.co.jp/jukyuu/
四国電力	http://www.yonden.co.jp/denkiyoho/download.html
九州電力	https://www.kyuden.co.jp/power_usages/pc.html
気象庁	http://www.data.jma.go.jp/obd/stats/etrn/index.php

- 実習用データの対象期間は**2016年4月1日～2017年12月31日**であり、**15,360時間分**のデータです。
  - 電力自由化の関係で2016年4月より一般電気事業者が提供するデータが変化したため、2016年4月を分析用データの期初としています。
  - このパートで利用する消費電力・気象データは、大規模標本のビックデータであるとともに改変を伴う利用や再配布ができるオープンデータでもあります。
- 分析作業を効率的に進めるために、下記のような**分析目的を設定**して、データ分析に取り組みます。

◆ 気象条件が消費電力に与える影響を分析し、気象条件の設定が与えられた場合に消費電力を予測したい。
◆ 寒すぎたり、暑すぎたりすると消費電力が上がることが見込まれるが、消費電力を最小化する気温を知りたい。

消費電力ビッグデータのRへの取り込み

◆実習用のExcelファイルを取り込んで、気象等が消費電力に与える影響を分析します。

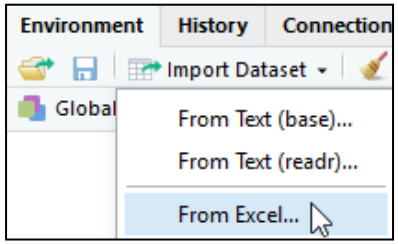
- 日時別の気象・消費電力データが格納されたExcelファイル「elec_weather.xlsx」をRへ取り込みます。
  - 様々な形態のデータを収集、整理統合することも、データ分析における重要なプロセスですが、ここでは省略して実習用のExcelファイルを取り込みます。
  - Excelファイルには9つの電力会社管内に関する消費電力（1時間単位：万kW）、気温（1時間単位：℃）、降水量（1日合計値：mm）のデータが並んでいます。なお、気象データは9つの一般電気事業者の本社所在地を対象地点としています。
  - 1時間単位の消費電力データは、その時点を期初とした後1時間のデータを示しますが、気象庁が公開する1時間単位の気象データは、その時点を期末とした前1時間のデータを示します。このため、Excelファイルでは消費電力データの表記に合わせて、公開された気象データを1時間前にずらしています。

実習用データ「elec_weather.xlsx」の内容（東京電力に関する列）

	A	B	C	D	E	G	H	K	T	AC
1	年月日時	年月	月	年月日	曜日	平日or土日祝	時	東京電力（1時間単位）	東京_気温（1時間単位）	東京_降水量（1日合計値）
2	2016/4/1 0時	201604	4	2016/4/1	金	平日	0	2555	13.8	2.5
3	2016/4/1 1時	201604	4	2016/4/1	金	平日	1	2433	13.0	2.5
4	2016/4/1 2時	201604	4	2016/4/1	金	平日	2	2393	12.2	2.5
5	2016/4/1 3時	201604	4	2016/4/1	金	平日	3	2375	11.2	2.5

- RStudioの[Import Dataset]ボタンまたは下記のコードによって、実習用データのExcelファイルを取り込みます。

実習用データ「elec_weather.xlsx」のRへの取り込み［入力コード05］




```
#1時間単位の消費電力、気象データのExcelファイルの取り込み
library(readxl)
elec_weather <- read_excel("C:/data/elec_weather.xlsx")
```

このスライドでは、実習用データを「C:/data/」下に設置している設定で説明します。

分析用データの抽出と変数名の改訂

◆分析に用いる変数を抽出し、プログラミングしやすいように変数名を英字に改訂します。

- Excelから取り込んだ全てのデータは、データフレーム「elec_weather」に入っていますが、日時のデータおよび回帰分析の対象とする地域のデータをデータフレーム「elwe」へと抽出します。
  - Rでは『新たなデータフレーム名<-抽出元のデータフレーム名[, 抽出したい変数名のベクトル] 』とコードを書くことによって、変数を抽出して新たなデータフレームを作成することができます。
- プログラミングをしやすいように**日本語の変数名を半角英数字の変数名へと変換**します。
  - Rでは日本語の変数名を取り扱えますが、プログラミングにおいては半角英数字の変数名が一般的であり、短い変数名の方が見やすいコードが書けます。
  - Rでは、『name(データフレーム名)<-c(“変数名1”, “変数名2”, …)』とコードを書くことによって、データフレームの左から新たな変数名を指定できます。

分析対象地域の電力に関するデータの抽出と変数名の改訂（東京電力のケース） ［入力コード06］

```
#日時データおよび東京電力データに関するデータの変数名のベクトルで指定
lm_name <- c("年月日", "月", "平日or土日祝", "時", "東京電力（1時間単位）", "東京_気温（1時間単位）" , "東京_降水量（1日合計値）")

#東京電力データに関する変数の列を抽出して、新たなデータフレームelweを作成
elwe <- elec_weather[, lm_name]

#プログラミングしやすいように変数名（列名）を英数字へ
names(elwe) <- c("ymd", "month", "hei_dns", "hour", "elec", "atemp" , "wdrop")
```

- 上記のコード例では、東京電力管内の消費電力に着目して、東京電力に関する変数を抽出しています。
  - プログラミングコードのファイルは地域別に9種を用意していますが、スライドでは東京電力管内に対応する「code_2_elec3tokyo.R」を用いて説明します。
  - 他の地域に関する分析を参照する場合は、コード内の（東京電力, 東京_）の部分を(北海道電力, 札幌_）（東北電力, 仙台_）（中部電力, 名古屋_）(北陸電力, 富山_）（関西電力, 大阪_）（中国電力, 広島_）（四国電力, 高松_）（九州電力, 福岡_）へと変更するか、各地域に該当するコードが記入されたファイルを参照してください。

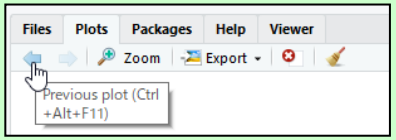
散布図による変数間の関係の確認

◆散布図によって、気象に関するデータと消費電力データの関係を確認します。

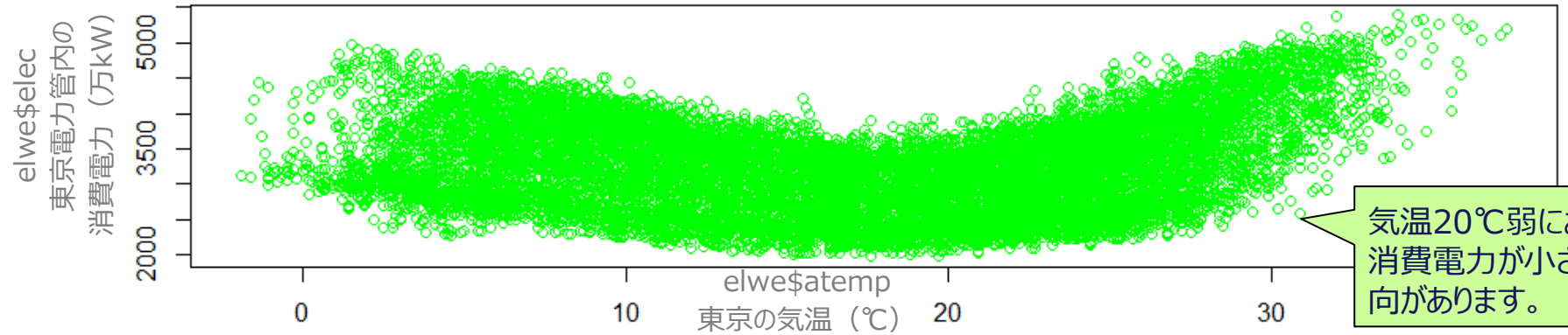
「気温と消費電力」「降水量と消費電力」の散布図の作成［入力コード07］

```
#「気温と消費電力(緑色)」「降水量と消費電力（青色）」のそれぞれの散布図の作成
plot(elwe$atemp, elwe$elec, col="green")
plot(elwe$wdrop, elwe$elec , col="blue")
```

RStudioでは、複数の図を出力しても、矢印ボタンで表示を移すことができます。

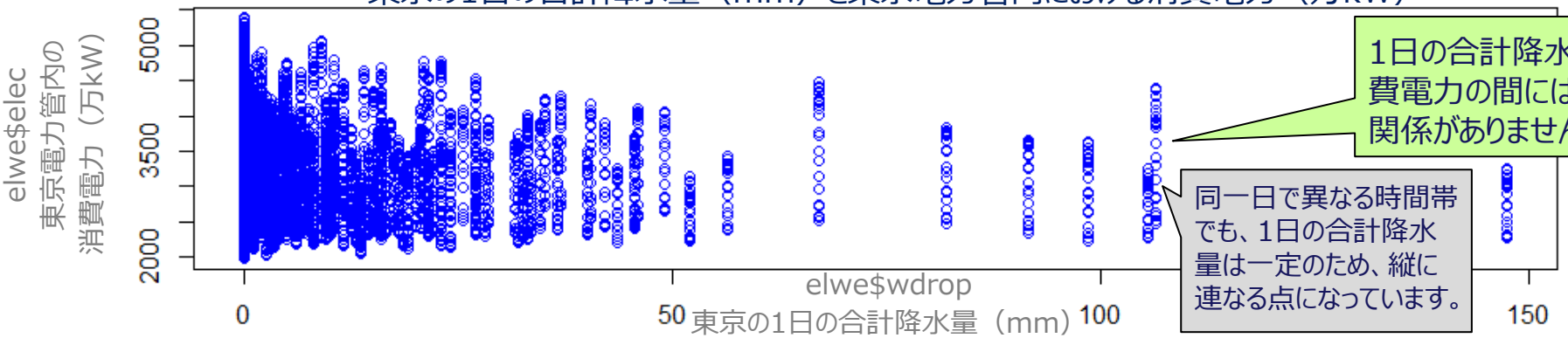


東京の気温（℃）と東京電力管内における消費電力（万kW）



気温20℃弱において、消費電力が小さくなる傾向があります。

東京の1日の合計降水量（mm）と東京電力管内における消費電力（万kW）



1日の合計降水量と消費電力の間には明瞭な関係がありません。

同一日で異なる時間帯でも、1日の合計降水量は一定のため、縦に連なる点になっています。

◘ 降水があると、気温が下がる傾向にあるため、散布図や単回帰分析では、各変数の正確な効果を把握できません。

累乗項・ダミー変数を説明変数とする重回帰分析

◆散布図を踏まえて、説明変数を設定して重回帰分析を行います。

- 気温と消費電力の非線形の関係を表すため、気温を2乗および3乗した値を説明変数に加えます。
  - 同一の説明変数を2乗すれば2乗項、3乗すれば3乗項となり、総じて累乗項といいます。
- 連続値としての降水量は、消費電力との明瞭な関係が見られないため、ダミー変数として「降水の有無」を設定します。
  - 該当する場合を1、該当しない場合を0と表し、有無や該当・非該当の状態を0と1で表した変数をダミー変数といいます。
  - Rにおけるifelse関数は、『ifelse([条件], [条件に合う場合の出力], [条件に合わない場合の出力])』と記述して、ダミー変数の作成に利用できます。
- 回帰分析の結果の推定係数を出力し、CSVファイルに保存します。
  - Rにおけるcoef関数は括弧内に回帰分析の結果データを指定することで、回帰分析の推定係数を表示できます。
  - Rにおけるwrite.table関数は、『write.table([R内の保存対象], "[PC内の出力先]", sep="[区切り文字]")』と記述して、ファイル保存ができます。

気候データを用いた回帰分析の実行と推定係数の表示と保存［入力コード08］

```
#地域の気温の2乗項、3乗項を作成する
elwe$atemp2<- elwe$atemp^2
elwe$atemp3<- elwe$atemp^3

#1日の降水有無のダミー変数を作成する
elwe$wadr_dummy<- ifelse(elwe$wdrop>0, 1, 0)

#地域の電力利用を気温（3乗項まで）と降水有無のダミー変数で回帰分析を行い、lm_re1へ結果データを保存
lm_re1<-lm(elwe$elec~elwe$atemp+elwe$atemp2+elwe$atemp3+elwe$wadr_dummy)

#回帰分析の結果データより推定係数を表示
coef(lm_re1)
#回帰分析の推定係数をCSVファイルに保存
write.table(coef(lm_re1), "C:/data/lm_re1_coef.csv", sep=",")
```

推定係数の一覧をコンマ区切りで「lm_re1_coef.csv」へ保存します。

「気温」と「降水の有無」による重回帰分析の結果表示

code_2_elec3tokyo.R
入力コード08の出力

◆重回帰分析の分析結果の推定係数を確認します。

● 消費電力を「3乗項までの気温の累乗項」「降水の有無のダミー変数」で回帰分析を行った結果の推定係数として、下記のような推定値が得られます。

```
> coef(lm_re1)
   (Intercept)       elwe$atemp      elwe$atemp2      elwe$atemp3 elwe$wadr_dummy
  3453.6410544       35.4280133       -8.5273962         0.2650495       68.5061220
```

● カンマで区切られたテキストファイルに当たるCSVファイルにおいても、推定係数を出力することができます。

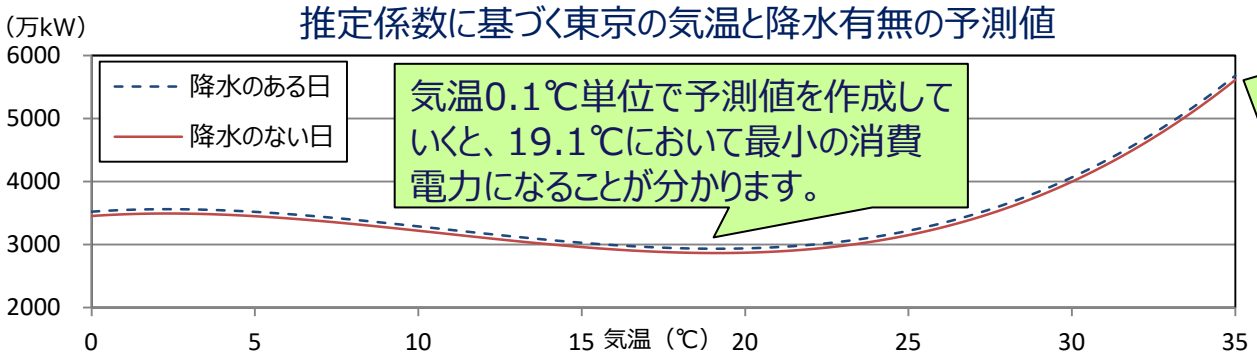
- CSVファイルは、Windowsのメモ帳のようなテキストエディターでも、Excelでも開くことができます。

```
"x"↓
"(Intercept)",3453.6410544404↓
"elwe$atemp",35.4280132539264↓
"elwe$atemp2",-8.52739622209195↓
"elwe$atemp3",0.265049482073631↓
"elwe$wadr_dummy",68.5061220158637↓
```

	A	B
1	x	
2	(Intercept)	3453.641
3	elwe$atemp	35.42801
4	elwe$atemp2	−8.5274
5	elwe$atemp3	0.265049
6	elwe$wadr_dummy	68.50612

● 推定係数に基づいて、各気温の当てはめ値を導出し最小の消費電力をもたらす気温を導出することができます。

- 『Intercept+atemp･気温+atemp2･気温^2+atemp3･気温^3+wadr_dummy･降水有無の設定』で気温ごと･降水別の予測値が得られます。



気温0.1℃単位で予測値を作成していくと、19.1℃において最小の消費電力になることが分かります。

「降水のある日」は「降水のない日」に比べて68.5万kW消費電力が大きくなる予測値がダミー変数の推定係数から示されます。

例示した回帰分析の式の形（ダミー変数の入れ方）では、1日の降水の有無によって予測値が上下に並行移動する形になっています。

月別・時間帯別の消費電力の箱ひげ図

◆月別・時間帯別の消費電力を箱ひげ図によって確認します。

● 月別・時間帯別の消費電力を見るために、データの種類を変換し、月別・時間帯別の**箱ひげ図**を描きます。

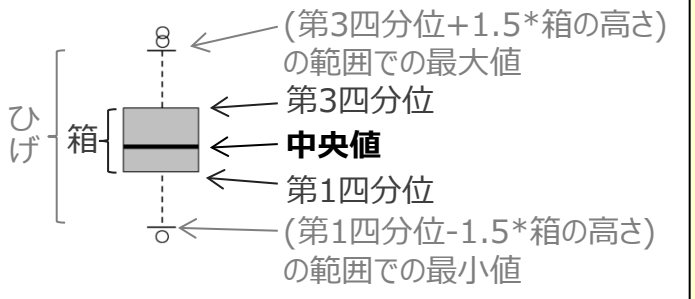
日付に関するデータの変換と箱ひげ図の作成［入力コード09］

```
#休日なら1、平日なら0のダミー変数を作成する
elwe$off_dummy<- ifelse(elwe$hei_dns=="土日祝", 1, 0)

#plot関数で箱ひげ図を描くためにも、
#後にダミー変数を作成するmakedummiesパッケージを利用するためにも
#月と時間帯のデータの型を因子型（factor）へ変更
elwe$month <- as.factor(elwe$month)
elwe$hour <- as.factor(elwe$hour)

#plot関数は因子型データを一つ目、連続データ二つ目に入れると箱ひげ図を表示
plot(elwe$month, elwe$elec ,col="green")
plot(elwe$hour, elwe$elec, col="cyan")
```

Ifelse関数内でとイコールを2つ連ねることで、文字列が等しい条件を表します。

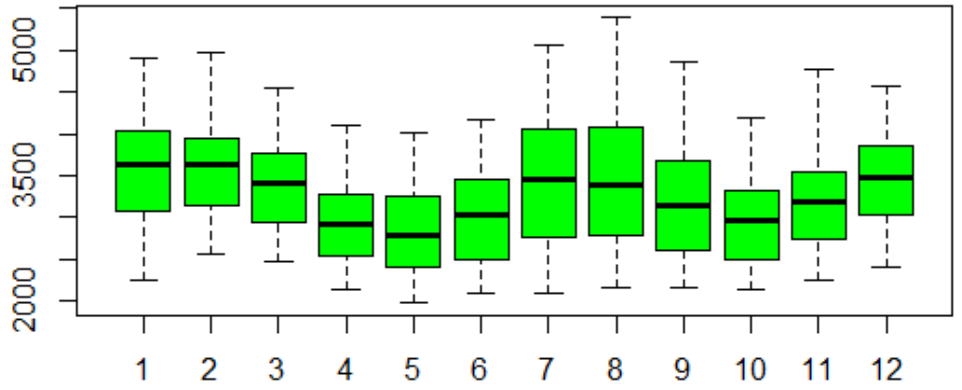
「因子型」とは、カテゴリーを表すデータの型を指しています。

Rのplotにおける箱ひげ図の表記

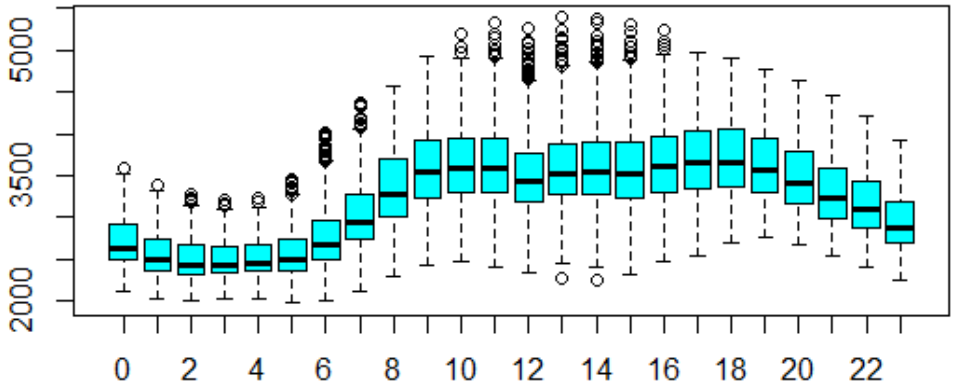


※ Rの箱ひげ図では「ひげ」の範囲外となる値を「外れ値」として、○で示します。

月別の消費電力の箱ひげ図



時間帯別の消費電力の箱ひげ図



□ 夏・冬の消費電力が春・秋に比べて大きい傾向、日中の消費電力が夜間に比べて大きい傾向が示されています。

月別・時間帯別のダミー変数の作成

◆回帰分析において説明変数として用いる　月別・時間帯別のダミー変数を作成します。

- makedummiesパッケージを利用して、月別・時間帯別のダミー変数を作ります。
  - makedummiesパッケージの利用において、初期値でもある『basal_level=FALSE』にすることで、最も小さい値（例：1月、0時～）のダミー変数を作成しません。一方、『basal_level=TRUE 』にすると、最も小さい値のダミー変数を含めて作成します。
  - 最も小さい値のダミー変数を含めて回帰分析を行うと、他の説明変数の定数倍と和が特定の説明変数と等しくなる「完全な多重共線性」という問題が生じ、回帰分析ができなくなります。Rのlm関数では「完全な多重共線性」があっても、当該説明変数を自動で外して回帰分析を行いますが、ここではあらかじめ、最も小さい値のダミー変数を作成しません。

makedummiesパッケージによるダミー変数の作成［入力コード10］

```
#install.packages("makedummies")
library(makedummies)

#makedummiesを使って月ダミーと時間帯ダミーを作成する
#「basal_level = FALSE」にすると、一番小さい値の1月および0時のダミー変数を作成しない。
md<-makedummies(elwe, basal_level = FALSE,  col = "month")
hd<-makedummies(elwe, basal_level = FALSE,  col = "hour")

#月ダミーと時間帯ダミーの統合
mdhd<-data.frame(md, hd)
```

- 環境（ワークスペース）に表示されたデータをクリックすることで、作成したダミー変数の内容を確認できます。
  - 『時間･月ダミー<-data.frame(月ダミー, 時間ダミー) 』と記述し、2種類のダミーデータを横に並べた時間･月ダミーのデータフレームを作成しています。

データをクリック

月別のダミー変数(md)

	month_2	month_3	month_4	month_5
1	0	0	1	0
2	0	0	1	0
3	0	0	1	0
4	0	0	1	0

時間帯別のダミー変数（hd）

	hour_1	hour_2	hour_3	hour_4
1	0	0	0	0
2	1	0	0	0
3	0	1	0	0
4	0	0	1	0

日時に関するダミー変数を含めた回帰分析の実行

code_2_elec3tokyo.R
入力コード11

◆休日・月別・時間帯別のダミー変数を含めて消費電力の回帰分析を行います。

- 日時に関するダミー変数を用いて、3種類の回帰分析を行います。
  - lm関数において、説明変数に『+.』を記入すれば、『, data= 』の後に記入したデータセットを全て説明変数に含めます。ここでは、makedummiesパッケージを用いて作成した月ダミーセットのmd、時間ダミーセットのhd、月・時間ダミーセットのmdhdをそれぞれ説明変数に加えます。
- これまでの回帰分析の結果をmemiscパッケージのmtable関数を使って比較できる形で出力します。
  - 下記のコードにおいては、R内の出力とCSVファイルに保存する方法をそれぞれ示しています。
  - mtableの初期設定では推定係数にP値に基づく「*」がついており、CSV等で数値としての取り扱いの支障になることもあります。「*」を除いた出力にするには、mtable関数内で『,signif.symbols=c("***"=-.001,"**"=-.01,"*"=-.05)』と初期設定の検定水準の値の前にマイナスを付けてください。

休日・月別・時間帯別のダミー変数を含めた回帰分析［入力コード11］

```
#lm関数において、説明変数に「+.」を記入すれば「, data=」の後のデータセットを全て説明変数に含める
#地域の電力利用を気温（3乗項まで）と降水ダミー、休日ダミー、月ダミー
lm_re2<-lm(elwe$elec~elwe$atemp+elwe$atemp2+elwe$atemp3+elwe$wadr_dummy+elwe$off_dummy+. ,data = md)

#地域の電力利用を気温（3乗項まで）と降水ダミー、休日ダミー、時間帯ダミー
lm_re3<-lm(elwe$elec~elwe$atemp+elwe$atemp2+elwe$atemp3+elwe$wadr_dummy+elwe$off_dummy+. ,data = hd)

#地域の電力利用を気温（3乗項まで）と降水ダミー、休日ダミー、月ダミー、時間帯ダミー
lm_re4<-lm(elwe$elec~elwe$atemp+elwe$atemp2+elwe$atemp3+elwe$wadr_dummy+elwe$off_dummy+. ,data = mdhd)

#install.packages("memisc")
library(memisc)
#回帰分析の結果を比較できる形でRの出力に表示
mtable(lm_re1, lm_re2, lm_re3, lm_re4)

#回帰分析の結果を比較できる形でCSVファイルに保存
lm_re1t4<-mtable(lm_re1, lm_re2, lm_re3, lm_re4)
write.mtable(lm_re1t4,file="C:/data/lm_re1t4.csv",colsep=",")
```

日時に関するダミー変数を含めた回帰分析の結果表

code_2_elec3tokyo.R
入力コード11の出力CSVの確認

◆日時に関するダミー変数を含めた回帰分析の結果の出力ファイルの内容を確認します。

● 回帰分析の出力が入ったCSVファイルをExcelなどで開いて、結果を確認します。

- ExcelでCSVファイルを直接開くと、半角括弧内にある各推定係数の標準誤差は負の値と見なされます。この問題を回避するには、Excelの「データ」タブの「テキストファイルの取り込み」から当該CSVファイルを指定し、「カンマ区切り」を設定し、「列のデータ形式」を全て「文字列」として読み込んでください。

回帰分析の結果の比較出力（抜粋）

		気象データのみの説明変数（基本形）	休日・月ダミーを説明変数に追加	休日・時間帯ダミーを説明変数に追加	休日・月・時間帯ダミーを説明変数に追加
		lm_re1	**lm_re2**	**lm_re3**	**lm_re4**
切片	(Intercept)	3453.641***	3567.479***	3666.940***	3575.807***
		(26.756)	(26.589)	(15.399)	(15.411)
気温	elwe$atemp	35.428***	56.734***	-82.870***	-52.765***
		(6.010)	(5.845)	(2.832)	(2.890)
気温の2乗項	elwe$atemp2	-8.527***	-5.532***	-0.953***	-1.670***
		(0.391)	(0.389)	(0.184)	(0.189)
気温の3乗項	elwe$atemp3	0.265***	0.182***	0.115***	0.109***
		(0.008)	(0.008)	(0.004)	(0.004)
降水日ダミー	elwe$wadr_dummy	68.506***	98.521***	43.777***	36.243***
		(9.077)	(8.286)	(4.172)	(4.033)
休日（土日祝）ダミー	elwe$off_dummy		-391.024***	-387.540***	-388.911***
			(7.902)	(4.125)	(3.822)
：	：	：	：	：	：
月ダミー（1月基準）	month_2		-70.789**		10.826
			(24.523)		(11.869)
：	：	：	：	：	：
時間帯ダミー（0時基準）	hour_1			-152.426***	-151.062***
				(13.392)	(12.383)
：	：	：	：	：	：
決定係数	R-squared	0.291	0.456	0.851	0.873
自由度調整済み決定係数	adj. R-squared	0.290	0.456	0.851	0.873
：					

複数の回帰分析の式の形で推定係数を確認することで、推定係数の安定性を把握できます。

休日（土日祝）は平日に比べて、約390万kW消費電力が少ないことが示されています。

回帰分析の決定係数を比較すると、どの説明変数（群）の説明力が高いかが分かります。

自由度（＝標本数-説明変数の種類数）が、1万以上などの大きな数であれば「自由度調整済み決定係数」と「決定係数」は僅差になります。

回帰分析の予測値の作成（for関数によるループ）

◆4種類の回帰分析の結果に基づいて、各気温に対する消費電力の予測値を作成します。

● 4種類の回帰分析の推定結果を比較するために、**各気温における予測値を導出**します。

- 1乗項のみの回帰分析では、説明変数の水準に依存せず、各説明変数1単位の増加が被説明変数に同一の効果があると仮定していますが、累乗項を含む回帰分析では、推定係数の直感的な解釈が困難なため、予測値のグラフ等によって効果を可視化して確認します。
- 説明変数として回帰分析に利用したデータに対する予測値は、predict関数を利用して『elwe$pr_lm4 <- predict(lm_re4)』で簡単に導出することができますが、仮想的な説明変数に対する予測値を導出するため、推定係数と仮想的な説明変数を用いた計算によって予測値を作成します。

● 4種類の回帰分析に基づく予測値を短いプログラムで作成するためにfor関数による**ループ**を利用します。

- プログラムにおいて指定した条件を満たすまで、繰り返し行われる処理をループといいます。Rにおけるfor関数では『for([変数名] in [変数の下限の自然数]：[変数の上限の自然数])』と記入すると、中括弧で示されたループを回る度に変数の値が下限から上限まで1ずつ増加していきます。
- 下記のコードでは『**rn**』を変数名として、coef_re『**rn:1～4**』という4種の推定係数のベクトルを作成後、0.1℃～35.0℃の気温に対応する予測値の変数hypo$pr_re『**rn:1～4**』を作成しています。

回帰分析の結果に基づく4種の予測値の作成［入力コード12］

```
#仮想的な予測値作成用に350行の新たなデータフレームを作り、行番号に対応する変数lineを入れる
hypo <- data.frame("line" = c(1:350))
#行番号lineに0.1を掛けて0.1℃刻みで気温設定の変数atemp（0.1～35.0℃）を作成
hypo$atemp<-hypo$line*0.1

#forでrnに1～4までの数字が順次入るループを作り、結果1～4に関する仮想的な気温に基づく予測値を作成
for(rn in 1:4){
#推定結果の1～4からそれぞれの推定係数のベクトル[coef_re番号]を作成
eval(parse(text=paste("coef_re",rn,"<-coef(lm_re",rn,")",sep="")))
#4種の推定係数のベクトルに基づき「（1月平日の0時台）降水なし日」における気温毎の電力消費の予測値を作成
eval(parse(text=paste("hypo$pr_re",rn,"<-coef_re",rn,"[1]+coef_re",rn,"[2]*hypo$atemp+
        coef_re",rn,"[3]*hypo$atemp^2+coef_re",rn,"[4]*hypo$atemp^3",sep="")))
                }
```

paste関数は、関数内の最後の『sep=』に続く引用符内を区切り値として、文字を連結します。この場合は、『sep=“”』とすることで、区切り値のない文字連結となっています。
for関数で変化させたい値以外の文字列は、引用符で囲って記入します。

parse関数では、文字列をRのコマンドとして解釈します。eval関数では、Rのコマンドを実行します。

日時ダミーを含めた回帰分析の予測値グラフ

◆作成したデータから各気温における消費電力の予測値をグラフに示します。

- matplot関数を利用すると、複数の列のデータを用いてグラフを作成することができます。

4種の回帰分析に基づく気温と消費電力の予測値のグラフ作成［入力コード13］

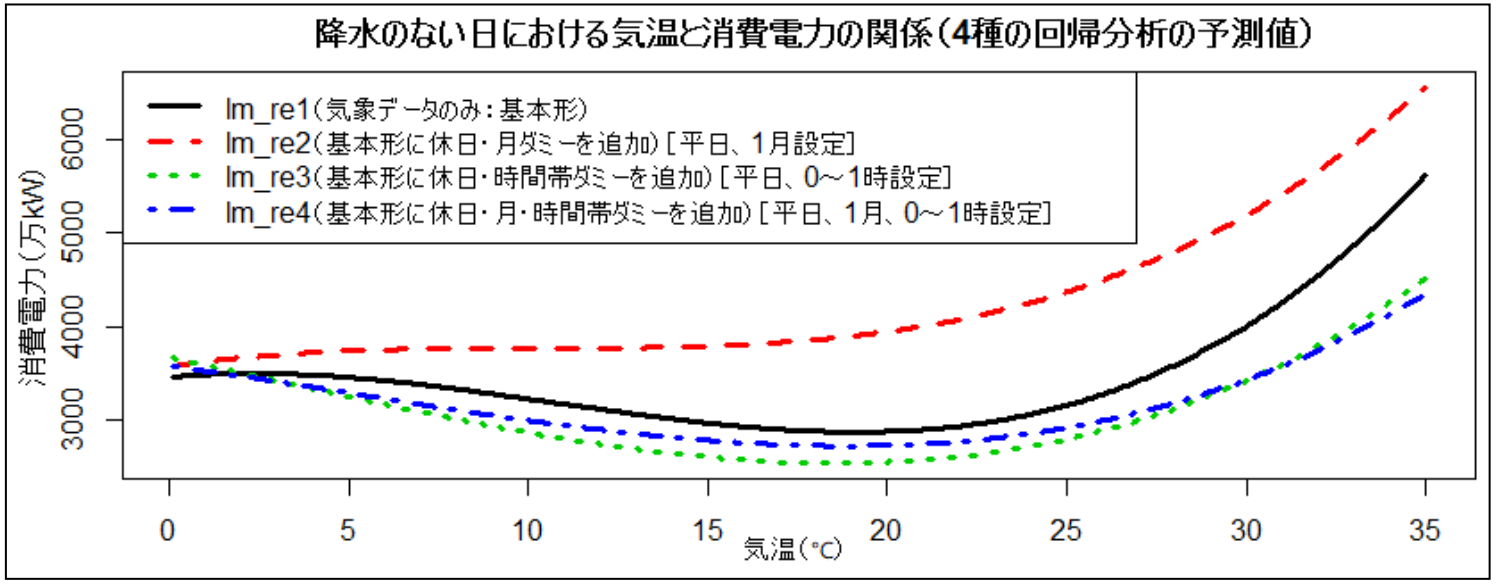
```
#気温毎の消費電力予測値のグラフ作成
matplot(hypo$atemp,hypo[ ,3:6], type="l",lwd = 3 , xlab="気温（℃）" , ylab="消費電力（万kW）")
title("降水のない日における気温と消費電力の関係（4種の回帰分析の予測値）")
legend("topleft", legend=c("lm_re1（気象データのみ：基本形）",
                           "lm_re2（基本形に休日・月ダミーを追加）［平日、1月設定］",
                           "lm_re3（基本形に休日・時間帯ダミーを追加）［平日、0～1時設定］",
                           "lm_re4（基本形に休日・月・時間帯ダミーを追加）［平日、1月、0～1時設定］")
                           ,col=1:4,lty=1:4,lwd = 3, text.width=25)
```

データフレーム「hypo」の3列目から6列目に入っている4種類の予測値を指定しています。

『type="l"』では、線によるグラフ作成を指定し、『lwd = 3』では線の太さを指定してます。

いったんグラフを作成した後に日本語でグラフのタイトル（title）と凡例（legend）を追記します。

Rにおける4種の回帰分析に基づく気温と消費電力の予測値のグラフ



回帰分析の結果に基づく測定・予測・発見

◆回帰分析の結果に基づいて定量的な効果の測定、予測、発見を行うことができます。

- 分析結果に基づいて、**説明変数の効果を定量的に測定すること**ができます。
  - 休日が平日に比べて消費電力が小さいことは経験的にも認知されていますが、回帰分析を行うことで東京電力管内では約390万kW小さくなることが定量的に測定できます。
- 回帰分析の推定結果を用いれば、月・時間帯の設定、天気予報における**消費電力の予測を行うこと**ができます。
  - 回帰分析（lm_re4）で得られた「気象データ、休日・月・時間帯ダミー」の結果を用いれば、「7月の平日14時において予測気温28.5℃、晴れ予報」といった任意の月・時間帯における天気予報に基づく消費電力の予測を行うことができます。
- 分析結果から、特定の条件を満たす値を導出したり、**新たな発見**が得られたりすることがあります
  - which.min関数では、最小値の変数があるデータフレームの位置を示すことができるため、変数の最小値やそれをもたらす条件を表示できます。回帰分析（lm_re4）に基づけば、東京電力管内において、最小の消費電力をもたらす気温は18.8℃であることが示されます。
  - 東京電力管内と同様の最小消費電力をもたらす気温を9つの電力管内で導出すると、日本の北東側で低く、南西側で高い傾向にあることが分かります。
  - コードの記述を少し改変するだけで、すぐに同種の処理・分析が行えることもプログラミングの利点の一つです。

休日・月・時間帯ダミーを追加した回帰分析で最小の消費電力をもたらす気温導出［入力コード14］

```
#lm_re4（基本形に休日・月・時間帯ダミーを追加）の予測値に基づく消費電力の最小値をもたらす気温
hypo[which.min(hypo$pr_re4),2]
```

データフレームhypoにおける「hypo$pr_re4」の最小値の行、2列目（気温の列）を表示します。

休日・月・時間帯ダミーを追加した回帰分析で最小の消費電力をもたらす気温

```
> hypo[which.min(hypo$pr_re4),2]
[1] 18.8
```

北海道電力	16.0℃	中部電力	19.7℃	中国電力	20.2℃
東北電力	17.5℃	北陸電力	18.3℃	四国電力	19.8℃
東京電力	18.8℃	関西電力	19.6℃	九州電力	19.4℃

□ 気温や日時データの他にも、追加を検討すべき説明変数として「電気代」が考えられます。本教材の【参考2】ではe-Stat APIから、「電力代」に関する物価指数を取得する手順を示しています。

【参考1】RとRStudioのダウンロード

◆RとRStudioは、誰でもウェブサイトからダウンロードすることができます。

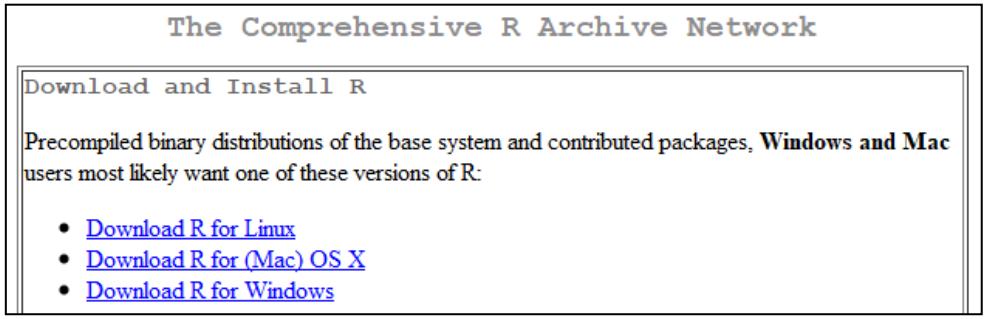
● Rのインストール用ファイルは、**CRAN**に参加する統計数理研究所のウェブサイトからダウンロードすることができます。

https://cran.ism.ac.jp/　　What are R and CRAN?

大学共同利用機関法人 情報・システム研究機構 統計数理研究所 The Institute of Statistical Mathematics

- CRAN（シーラン：Comprehensive R Archive Network ）は、Rに関するファイルを蓄積・提供する国際ネットワークです。
- 2018年3月時点における上記URLのウェブサイトの表記は概ね英語ですが、英単語が分かれば、ダウンロードやインストールに支障はありません。
- OSへインストールするためのRには、Windows版、Macintosh版、Linux版がありますが、この講座ではWindows版で説明します。
- Windowsを利用している場合は「Download R for Windows」をクリックした後に表示されるWindows版のダウンロードボタンをクリックしてください。

OSに応じたRの選択画面



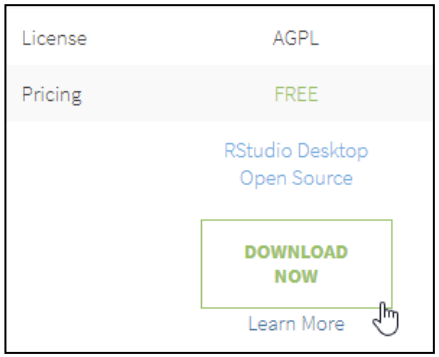
Windows用Rのダウンロード画面

● RStudioのインストール用ファイルは、RStudioのウェブサイトからダウンロードできます。

https://www.rstudio.com/products/rstudio/download/

デスクトップ版のダウンロードへのリンク

オープンソース版のダウンロードボタン



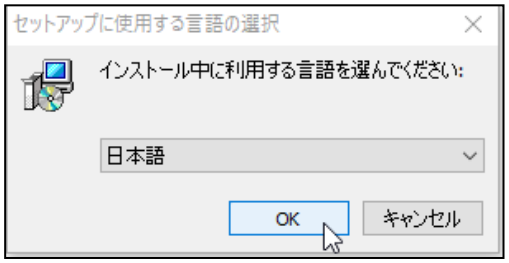
- RStudioには、各PCの中のRを実行するデスクトップ版と離れたサーバ上のRを実行するサーバ版がありますが、一般にはデスクトップ版を利用します。
- RStudioのトップページからの移動する場合は、まず画面上部の「Products＞RStudio」を選択してください。次に表示される画面で[Open Source Edition]の欄にある「DOWNLOAD RSTUDIO DESKTOP」のボタンを押します。続いて表示される画面でオープンソース版の「DOWNLOAD NOW」をクリックしてください。

【参考1】RとRStudioのインストール

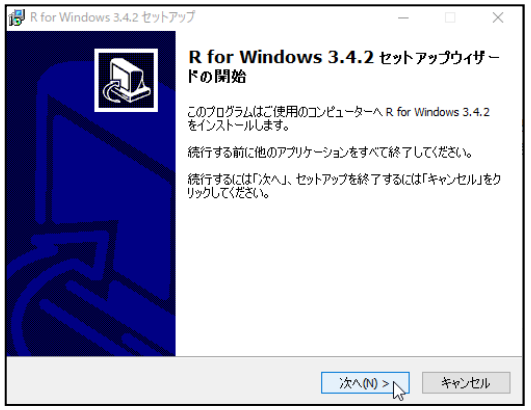
◆RとRStudioは、マウスによる操作だけで簡単にインストールすることができます。

- Rのインストールにおいては、全て初期設定で「OK」や「次へ」で進めて、問題ありません。
  - イントール時の初期設定が把握でき、変更したい方は、インストール先のフォルダの指定、32bit版か64bit版等の選択をしてください。設定内容が把握できない方や細かい設定を気にしない方は、全て初期設定でのインストール、32bit版と64bit版の両方のインストールで構いません。

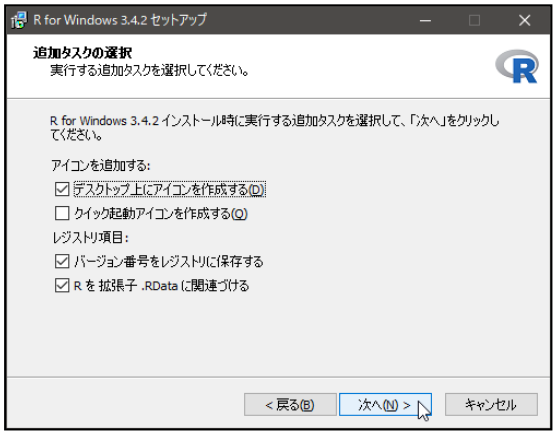
Rのインストールの言語選択



Rのインストール開始画面

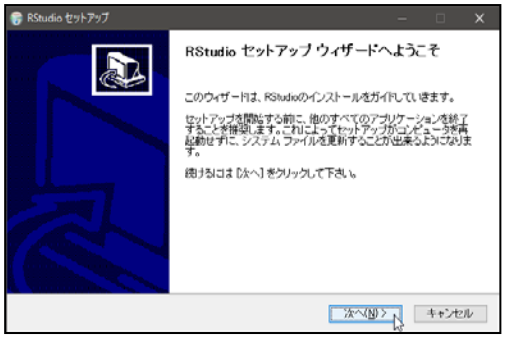


Rのインストール時の最後の選択

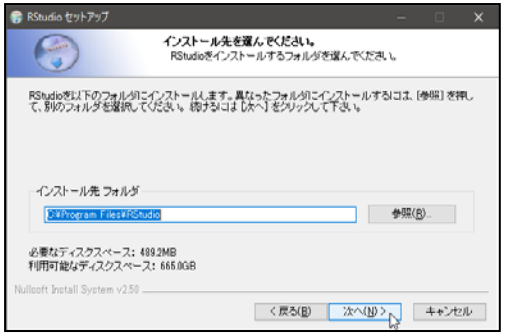


- RStudioのインストールも、全て初期設定で「次へ」で進めて、問題ありません。
  - 初期設定でインストールを完了すると、スタートメニューの中にRStudioのショートカットができます。

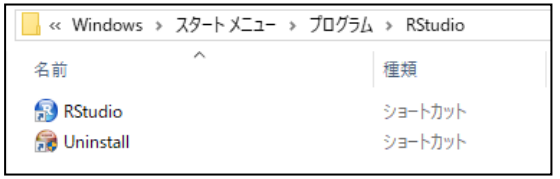
RStudioインストール開始



RStudioのインストール先指定



プログラムフォルダ内のショートカット

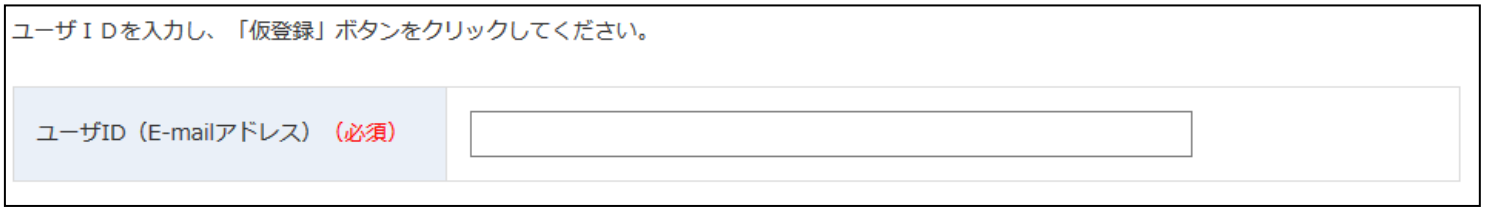


【参考2】e-Statへのユーザー登録

◆e-Stat APIを利用するために、まずe-Statにユーザー登録を行います。

- e-Statウェブサイトの「新規登録」または下記のURLからユーザー登録を行います。
  https://www.e-stat.go.jp/mypage/user/preregister
- ユーザーIDとなるE-mailアドレスを登録して、e-Statへの仮登録を行ってください。
  - 仮登録画面でメールアドレスを記入すると、そのメールアドレスに対してe-Statから自動送信メールが送付されます。そのメールの本文内にあるURLから本登録用のウェブページを開いてください。

e-StatへのユーザーID（E-mailアドレス）の仮登録欄

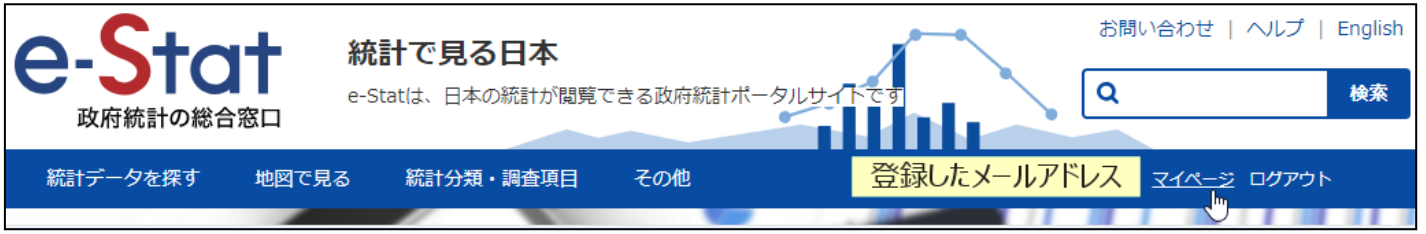


- 本登録においては、利用する機能に「API機能」を含めて指定し、ログイン時に利用するパスワードを指定してください。

e-Statのユーザー登録において利用する機能の選択

- 本登録の完了後e-Statにログインし、E-mailアドレス表示の右側にある「マイページ」をクリックしてください。

e-Statにログインした上で「マイページ」のクリック



【参考2】e-Stat APIのアプリケーションID発行

◆e-Stat APIを利用するためのアプリケーションIDの発行を行います。

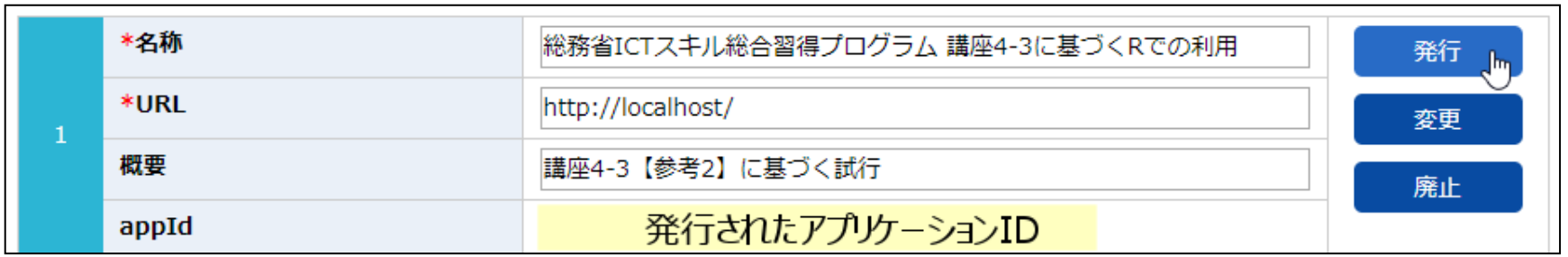
● マイページの上部に表示される「API機能（アプリケーションID発行）」をクリックしてください。

e-Statの「マイページ」内の「API機能（アプリケーションID発行）」のクリック

● 「名称」および「URL」を記入して、「発行」ボタンをクリックして、e-Stat APIのアプリケーションIDを発行します。

- URLは、ウェブサービスからe-Stat APIを利用する場合は当該ウェブサービスのURLを入力します。今回のようにRのプログラミングによってPCから利用したり、非公開サイトでの利用を行う場合は、「http://localhost/」と入力してください。

e-Statにおける「API機能（アプリケーションID発行）」の発行と表示



● プログラミングでは、表示されたアプリケーションIDを含めてコードを書き、e-Stat APIに情報提供を要求します。

- e-Stat APIのアプリケーションIDは英数字の羅列で構成されています。
- e-Stat APIのアプリケーションIDは、各利用者の各サービスに対応する形で発行されています。アプリケーションIDを他者に教えたり、公開したりすることがないように取り扱いには注意してください。

【参考2】Rにおけるestatapiパッケージの利用

code_ref2_estat_api.R
入力コード15

◆RではE-StatのAPIを簡単に利用できるestatapiパッケージを利用することができます。

- Rにおけるestatpiパッケージを利用して、e-Stat内で『消費者物価指数』が含まれる統計を検索します。
  - e-Statは総務省統計局が運営している国のサービスですが、Rのestatapiパッケージは有志の個人が作成したものです。estatapiパッケージの詳細な利用方法は、CRAN内の説明ファイル等を参照してください。（ftp://cran.r-project.org/pub/R/web/packages/estatapi/README.html）
  - e-StatのAPIを利用するプログラミングコードとして、実習用のデータ・コードに含まれる「code_ref2_estat_api.R」を用いて説明します。
  - 本教材における消費電力ビッグデータの分析と関連して、インターネットを経由して消費者物価指数の「電気代」のデータを取得することを想定します。

estatapiパッケージの利用宣言と統計名の検索［入力コード15］

```
#estatapiパッケージのダウンロードと利用宣言
#install.packages("estatapi")
library(estatapi)
yourID <- "各利用者が取得したe-Stat APIのアプリケーションIDを記入"

#e-Stat内を『消費者物価指数』で検索
estat_getStatsList(appId = yourID, searchWord = "消費者物価指数")
```

```
#データ形式tibbleにおいて、表示幅を制限しない設定
options(tibble.width = Inf)

#データ形式tibbleにおいて、全ての行を表示する設定
options(tibble.print_max = Inf)
```

この枠内のコードは記入しなくても、データの取得は可能ですが、Rの出力から各種IDを正確に確認するために必要な設定です。

- 実行結果として、『消費者物価指数』が含まれる公的統計が下記のように表示されます。

Rの出力における『消費者物価指数』が含まれる統計情報の表示

```
> #e-Stat内を『消費者物価指数』で検索
> estat_getStatsList(appId = yourID, searchWord = "消費者物価指数")
# A tibble: 4 x 13
  `@id`      STAT_NAME          GOV_ORG STATISTICS_NAME
  <chr>      <chr>              <chr>   <chr>
1 0000010212 社会・人口統計体系 総務省  都道府県データ 社会生活統計指標
2 0002050001 消費者物価指数     総務省  平成17年基準消費者物価指数
3 0003036792 消費者物価指数     総務省  平成22年基準消費者物価指数
4 0003143513 消費者物価指数     総務省  2015年基準消費者物価指数
```

取得対象とする『2015年基準消費者物価指数』の統計IDは「003143513」であることを確認します。

- 本教材においては、統計IDが「003143513」の『2015年基準消費者物価指数』を採りあげます。

【参考2】estatapiパッケージを用いたコードの確認

code_ref2_estat_api.R
入力コード16

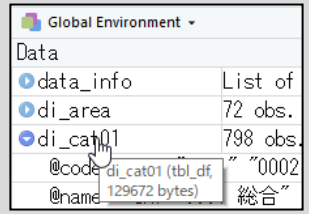
◆estatapiパッケージを用いて、「電気代」の地域別物価指数のコードを確認します。

- 統計IDが「003143513」の『2015年基準消費者物価指数』を指定し、データ説明を「data_info」に格納します。
- RStudioの環境内の「data_info」をクリックし、[cat01]に品目コード、 [area]に地域コードがあることを特定します。
- [cat01]と[area]を改めてRのデータに格納し、環境から変数をダブルクリックし、データビューでコードを確認します。

データ説明の格納と収集対象とするコードの確認［入力コード16］

```
#2015年基準消費者物価指数（statsDataId=0003143513）のデータ説明をdata_infoへ格納
data_info <- estat_getMetaInfo(appId=yourID, statsDataId="0003143513")

#品目コードと地域コードを調べるため、cat01およびareaをRのデータフレームとして格納
di_cat01<-data_info$cat01
di_area<-data_info$area
```

環境（ワークスペース）内の変数をダブルクリックして、データビューに表示します。



RStudioの環境（ワークスペース）内の表示

```
data_info          List of 5
 tab :Classes 'tbl_df', 'tbl' and 'data.frame': 5 obs. of
 ..$ @code : chr [1:5] "1" "2" "3" "4" ...
 ..$ @name : chr [1:5] "指数" "前月比・前年比・前年度比" "前年
 ..$ @level: chr [1:5] "" "" "" "" ...
 ..$ @unit : chr [1:5] NA "%" "%" NA ...
 cat01 :Classes 'tbl_df', 'tbl' and 'data.frame': 798 obs
 ..$ @code : chr [1:798] "0001" "0002" "0003" "0004" ...
 ..$ @name : chr [1:798] "0001 総合" "0002 食料" "0003 穀類" "
 ..$ @level : chr [1:798] "1" "1" "3" "5" ...
 ..$ @parentCode: chr [1:798] NA NA "0002" "0003" ...
 area :Classes 'tbl_df', 'tbl' and 'data.frame': 72 obs. o
 ..$ @code : chr [1:72] "13A01" "00000" "00011" "00012" ...
 ..$ @name : chr [1:72] "13100 東京都区部" "全国" "人口５万以上
```

物価指数の種類の変数名は［tab］です。

対象品目の変数名は［cat01］です。

対象地域の変数名は［area］です。

データビューでのコードの確認

	@code	@name
307	3185	3185 壁紙張替費
308	3180	3180 火災・地震保険料
309	0054	0054 光熱・水道
310	0056	0056 電気代

［di_cat01］より「電気代」の品目コード［0056］を確認します。

	@code	@name
22	01A01	01100 札幌市
23	02A01	02201 青森市
24	03A01	03201 盛岡市
25	04A01	04100 仙台市

［di_area］より「札幌市」の地域コード［01A01］を確認します。

【参考2】estatapiパッケージを用いたデータ取得

code_ref2_estat_api.R
入力コード17

◆estatapiパッケージを用いて、指定地域の「電気代」の物価指数を取得します。

- 「統計ID」「品目コード」「地域コード」等を指定して、各地域の「電気代」物価指数をepdataに格納します。
  - APIで指定可能なパラメーターは、e-Stat APIの仕様（https://www.e-stat.go.jp/api/e-stat-manual）を確認してください。

コード指定によるデータを取得とCSV保存［入力コード17］

```
#取得したコードを指定してe-Stat APIからデータを取得し、epdataに格納
epdata<-estat_getStatsData(appId = yourID,
  statsDataId = "0003143513", #2015年基準消費者物価指数
  cdTab="1", #物価指数（データ説明の変数名に接頭語としてcdを付け、変数名の最初を大文字で記入）
  cdCat01 = "0056" , #電気代
  cdArea = c("01A01","04A01","13A01","23A01","16A01","27A01","34A01","37A01","40A02","47A01"),
#札幌市、仙台市、・・・、福岡市の9地域指定
  lvTime ="4", #月次データ
  cdTimeFrom="2016") #データ取得開始年

#epdataをCSVに出力
write.csv(epdata, "C:/data/estat_elec_9area_price.csv")
```

あらかじめ、Cの直下に「data」というフォルダを作っておけば、その中にCSVファイル「estat_elec_9area_price.csv」が出力されます。

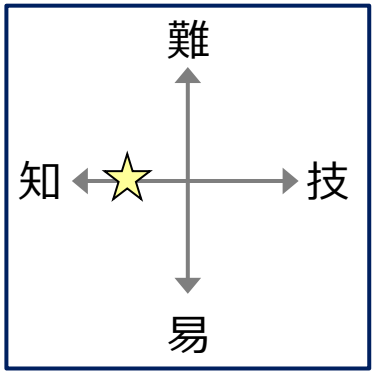
- epdataには指定した地域・時期のデータが格納されるとともに、データをCSVファイルに保存しています。

RStudioのデータビューにおけるepdataの表示

tab_code	表章項目	cat01_code	2015年基準品目	area_code	地域（2015年基準）	time_code	時間軸（年・月）	value
1	指数	0056	0056 電気代	01A01	01100 札幌市	2018000202	2018年2月	101.7
1	指数	0056	0056 電気代	01A01	01100 札幌市	2018000101	2018年1月	100.9

□ APIを利用すると、様々なデータソースから対象データを絞って収集することが可能となります。

総務省　ICTスキル総合習得教材



［コース4］オープンデータ・ビッグデータ利活用事例

4-4：分野別のICT・データの利活用事例

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_4_4.pdf

	1	2	3	4	5
［コース1］データ収集					
［コース2］データ蓄積					
［コース3］データ分析					
［コース4］データ利活用				●	

本講座の学習内容［4-4：分野別のICT･データの利活用事例］

【講座概要】

- 「官民データ活用推進基本法」とそれに基づく「官民データ活用推進基本計画」を示します。
- 「官民データ活用推進基本計画」における8つの重点分野のICT･データ活用事例を示します。
- 金融分野においては、ブロックチェーン技術の概要を示し、仮想通貨をはじめとするブロックチェーンの利用例を示します。
- 移動分野においては、段階的な自動運転の進展および市場化が期待される時期を紹介します。

【講座構成】

座学

官民データ活用推進基本法と官民データ活用推進基本計画	
(1)電子行政	(2)健康･医療･介護
(3)観光	(4)金融
(5)農林水産	(6)ものづくり
(7)インフラ･防災･減災	(8)移動

【学習のゴール】

- ✓ 官民データ活用推進基本法と官民データ活用推進基本計画の要点を理解する。
- ✓ 官民データ活用推進基本計画における8つの重点分野のICT･データ活用事例を把握する。
- ✓ ブロックチェーン技術の利用例を紹介できる。
- ✓ 段階的な自動運転の進展を把握する。

官民データ活用推進基本法

◆2016年12月に「官民データ活用推進基本法」が施行されました。

- 2016年12月に施行された**官民データ活用推進基本法**は、官民のデータ活用を推進するための法律です。

第1条　この法律は、（中略）官民データ活用の推進に関する施策を総合的かつ効果的に推進し、もっと国民が安全で安心して暮らせる社会及び快適な生活環境の実現に寄与することを目的とする。

官民データ活用推進基本法の基本的施策（10条～19条）の各条文の内容

条	条文の内容	条	条文の内容
10条	行政手続等のオンライン化原則	15条	情報システム改革
11条	オープンデータの推進	16条	IoT・AIなどの研究開発・実証の推進
12条	データ利活用のルール整備	17条	データ活用ができる人材育成
13条	マイナンバーカードの普及・活用	18条	データ活用の普及啓発
14条	デジタルデバイド対策	19条	国と地方の施策の整合性

【出所】「官民データ活用推進戦略会議の公表資料［首相官邸］」に基づき作成
https://www.kantei.go.jp/jp/singi/it2/dai71/gijisidai.html

- 官民データ活用推進基本法では、日本政府が**官民データ活用推進基本計画**を定めることを義務づけています。

第8条　政府は、官民データ活用の推進に関する施策の総合的かつ効果的な推進を図るため、官民データ活用の推進に関する基本的な計画（以下「官民データ活用推進基本計画」という。）を定めなければならない。

- 官民データ活用推進基本法第9条では、都道府県に「都道府県官民データ活用推進計画」を定めることを義務づけ、市町村（特別区を含む）に「市町村官民データ活用推進計画」を定める努力を求めています。

【出所】官民データ活用推進基本法［e-Gov|総務省］
http://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/detail?lawId=428AC1000000103

□ 日本政府や都道府県はそれぞれの「官民データ活用推進計画」を定めることが法律に定められています。

世界最先端ＩＴ国家創造宣言・官民データ活用推進基本計画

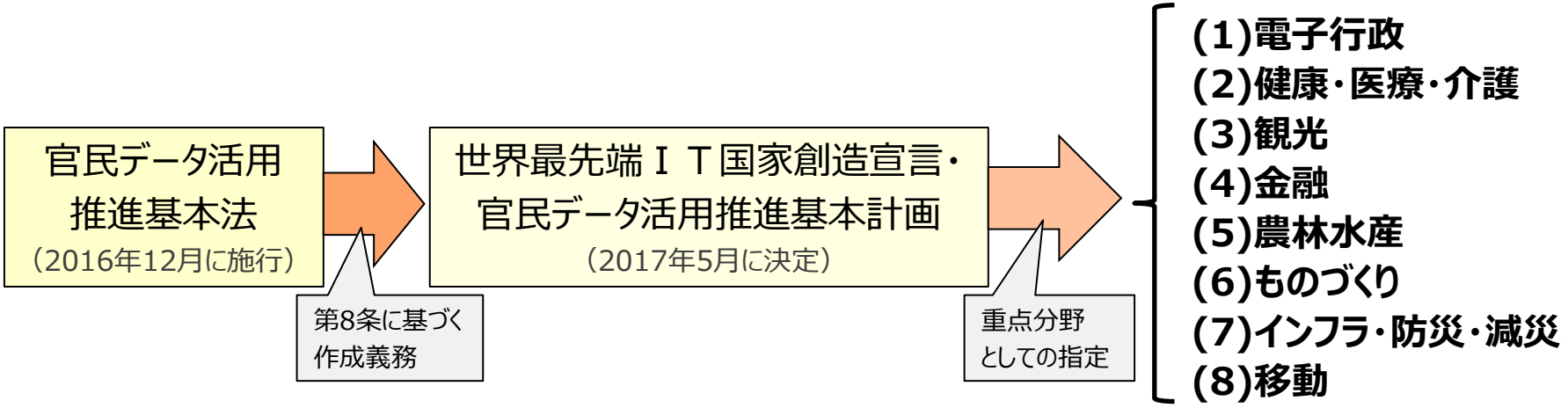
◆「世界最先端ＩＴ国家創造宣言・官民データ活用推進基本計画」が閣議決定されました。

- 2017年5月に**世界最先端ＩＴ国家創造宣言・官民データ活用推進基本計画**が閣議決定されました。
  - 本教材では主にICT（情報通信技術）という言葉を使っていますが、ICTとIT（情報技術）は、概ね同じ意味で使われています。
  - 「世界最先端ＩＴ国家創造宣言」は、2013年6月に初めて閣議決定された世界最高水準のIT活用社会の実現を目指す宣言です。
  - 「世界最先端ＩＴ国家創造宣言・官民データ活用推進基本計画」は、「世界最先端ＩＴ国家創造宣言」と「官民データ活用推進基本計画」を一体化する形で策定したIT・データの利活用を推進する宣言・計画です。

【出所】世界最先端IT国家創造宣言・官民データ活用推進基本計画［高度情報通信ネットワーク社会推進戦略本部（ＩＴ総合戦略本部）］
http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20170530/siryou1.pdf

- 「世界最先端ＩＴ国家創造宣言・官民データ活用推進基本計画」では**８つの重点分野**［**(1)電子行政**、**(2)健康・医療・介護**、**(3)観光**、**(4)金融**、**(5)農林水産**、**(6)ものづくり**、**(7)インフラ・防災・減災**、**(8)移動**］が示されています。

世界最先端ＩＴ国家創造宣言・官民データ活用推進基本計画における8つの重点分野



□ 本講座では、これらの8つの重点分野におけるICT（情報通信技術）の活用を紹介します。

(1)電子行政（行政手続のオンライン化）

◆日本政府の電子行政の取り組みとして、行政手続のオンライン化を推進しています。

- 日本政府の電子行政の取り組みの一環として、様々な**行政手続のオンライン化**が推進されています。

オンラインによる行政手続が可能な主なシステム

システム名・ＵＲＬ	主な手続
登記・供託オンライン申請システム http://www.touki-kyoutaku-online.moj.go.jp/	登記申請、登記事項証明書等の交付請求、供託申請等
国税電子申告・納税システム（e-Tax） http://www.e-tax.nta.go.jp/	国税申告手続、給与所得の源泉徴収票（及び同合計表）、国税納付手続等
電子政府の総合窓口（e-Gov） http://www.e-gov.go.jp/shinsei/	厚生年金保険、健康保険、雇用保険等社会保険・労働保険関係分野の手続等
輸出入・港湾関連情報処理システム（NACCS） http://www.naccs.jp/	輸入（納税）申告、輸出申告、船舶等の入出港手続等
特許庁電子出願システム http://www.pcinfo.jpo.go.jp/site/	特許庁に対する産業財産権出願関連手続（特許・実用新案・意匠・商標の出願手続及び登録料の納付手続等）
自動車保有関係手続のワンストップサービス（OSS） http://www.oss.mlit.go.jp/portal/	自動車の新車新規登録、変更登録、移転登録、抹消登録、継続検査

【出所】電子政府利用促進週間のお知らせ［e-Gov|総務省］
http://www.e-gov.go.jp/news/egov/2017/news20171023.html

- 行政手続のオンライン化によって、国民の利便性が向上するとともに、行政部門の効率化が図られます。
  - オンラインで申請された電子情報は、RPA（Robotic Process Automation）と呼ばれるソフトウェアロボットによる自動処理および人工知能による自動処理との親和性が高く、人手をかけずに処理することができるようになります。

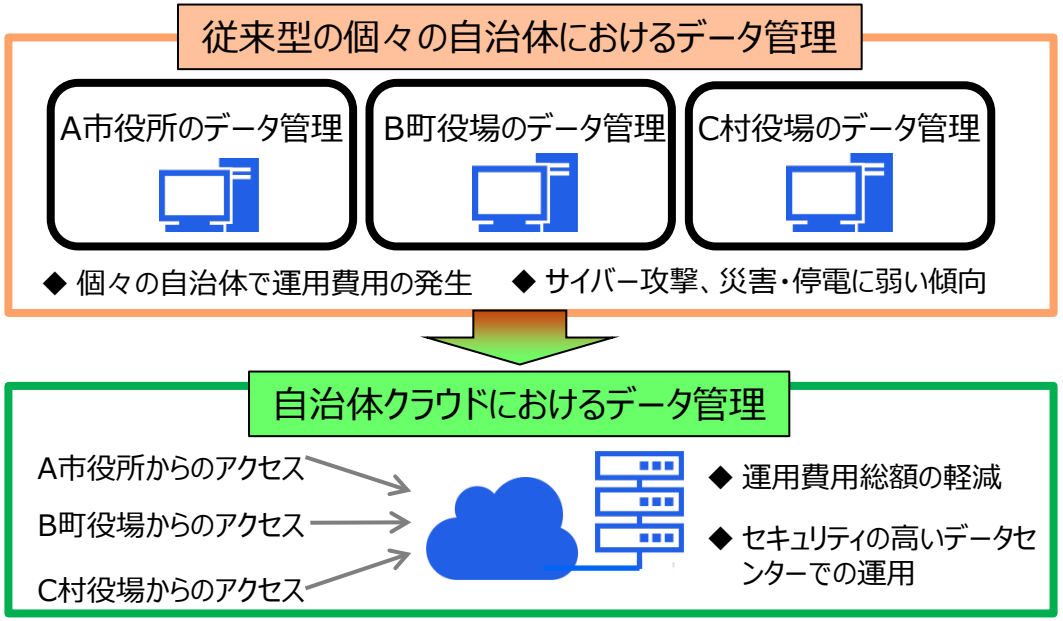
(1)電子行政（自治体に関する取り組み）

◆電子行政の推進は、地方自治の効率化、住民の利便性向上にもつながっています。

- 地方自治体が、情報システム・データを庁舎外のデータセンターで保有・管理し、ネットワークを通じて利用する**自治体クラウド**の取り組みも進展しています。
- 自治体クラウドを利用することで、自治体のデータシステムの運営費用総額を圧縮するとともにセキュリティを高めることが可能です。

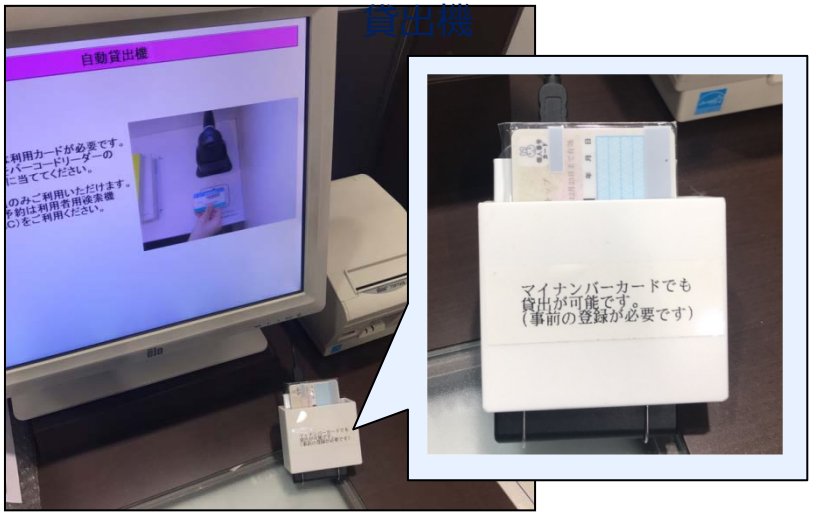
- 豊島区立中央図書館では、2017年9月に**マイナンバーカードによる図書の貸し出しを可能とする**図書館共同利用システム実証実験を開始しました。
  - 豊島区立中央図書館における図書館共同利用システム実証実験では、2017年9月に国がクラウド型共同利用システムとして運用開始した「マイキープラットフォーム」を活用することで、将来的に複数自治体の図書館利用カードをマイナンバーカード1枚に集約できる様式となっています。

従来型のデータ管理から自治体クラウドへの変化のイメージ



【出所】自治体クラウドポータルサイト［総務省］の情報に基づき作成
http://www.soumu.go.jp/main_sosiki/jichi_gyousei/c-gyousei/lg-cloud/

マイナンバーカード利用時のプライバシーに配慮した自動貸出機



【出所】全国初、マイナンバーカードを活用した「図書館共同利用システム」実証デモンストレーション［豊島区］
https://www.city.toshima.lg.jp/013/kuse/koho/hodo/h2909/1709121058.html

(2)健康・医療・介護（保険医療分野におけるAI活用）

◆医療におけるAIの活用は、様々な領域で検討されています。

- 厚生労働省では、2017年に**「保健医療分野におけるAI活用推進懇談会」**を開催し、有識者を招いて医療分野におけるAIの活用と開発に向けた施策を議論し、下記の方針をとりまとめました。

保健医療分野におけるAI活用領域と開発に向けた施策

分類	領域	AIの開発に向けた施策
AIの実用化が比較的早いと考えられる領域	ゲノム医療	• 実用化まで最も近いのは『がん』であり、実現に向けた推進体制を構築
	画像診断支援	• 病理・放射線・内視鏡等について、国内には質の高いデータが大量に存在しており、効率的な収集体制の確立が必要 ⇒ ・ 関連学会が連携して画像データベースを構築 ・ AIの開発をしやすくするため、薬事審査の評価指標の策定や評価体制の整備も実施
	診断・治療支援 (問診や一般的検査等)	• AIの開発をしやすくするため、医師法上や医薬品医療機器法上の取扱を明確化 • 各種データベース（ゲノム解析データを含む）の集約等により、難病を幅広くカバーする情報基盤を構築し、AIの開発に活用
	医薬品開発	• 健康医療分野以外でもAI人材は不足しているため、効率的なAI開発が必要（IT全体で30万人不足、うちAIで5万人不足）であり、製薬企業でもAI人材が不足 ⇒AI人材の有効活用の観点から、製薬企業とIT企業のマッチングを支援
AIの実用化に向けて段階的に取り組むべきと考えられる領域	介護・認知症	• 現場のニーズに基づかずに開発されたAI（技術指向のAI）では、現場には普及せず ⇒介護現場のニーズを明確化し、ニーズに基づく研究開発を実施
	手術支援	• 手術時のデジタル化データ（心拍数、脳波、術野画像等）は相互に連結されていない状態で、手術行為と各種データがリンクせず、AIによる学習が困難 ⇒手術関連データを相互に連結するためのインターフェースの標準化を実施

【出所】「保健医療分野におけるAI活用推進懇談会報告書 概要［厚生労働省］」に基づき作成
http://www.mhlw.go.jp/stf/shingi2/0000169233.html

(2)健康・医療・介護（介護分野）

◆IoTを活用することで、介護負担の軽減や異常の早期発見が可能となります。

- 介護分野では**介護される人のプライバシーに配慮してカメラを利用しないサービス**があります。

- パラマウントベッド株式会社が発売する「眠りSCAN」は、マットレスの下に設置したセンサーより、ベッドへの着床や睡眠状態（寝返り、呼吸、心拍）を把握します。

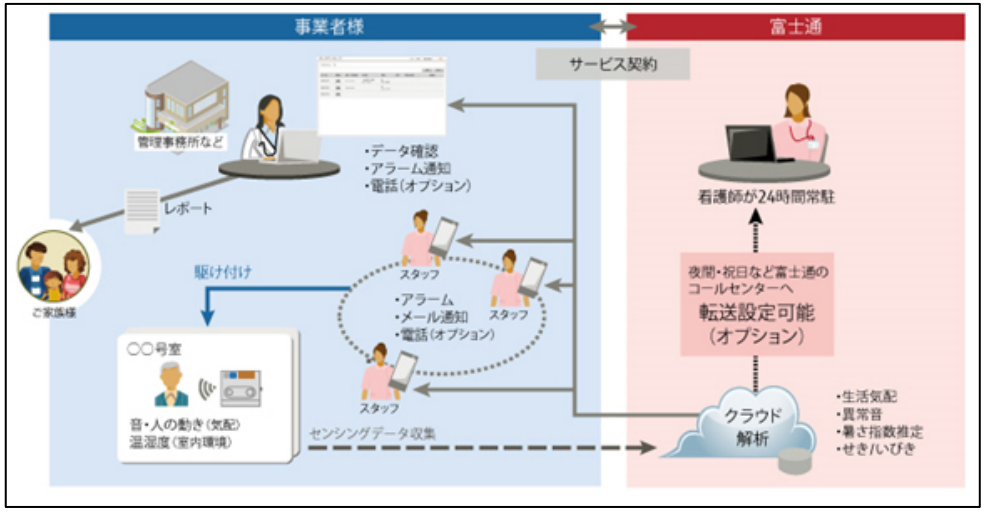
- 介護者はPC等で、施設の各入居者のベッドへの着床や睡眠状態をリアルタイムに一覧することができ、異常の早期発見や介護者の負担軽減に役立ちます。

- 富士通株式会社が2018年1月より提供している「リモートモニタリングサービス」では、部屋内に設置した音響センサーから得られた生活音をクラウドにて分析し、転倒の可能性など異常が疑われる場合にスタッフの駆けつけや家族への連絡を行うサービスです。

眠りSCANのリアルタイムモニター機能

【出所】見守り支援システム　眠りSCAN［パラマウントベッド株式会社］
http://www.paramount.co.jp/learn/reductionworkburden/nemuriscan

音響センサーによるリモートモニタリングサービス



【出所】「リモートモニタリングサービス」を提供開始［富士通株式会社］
http://pr.fujitsu.com/jp/news/2018/01/9.html

(3)観光（コミュニケーションロボット・ARの活用）

◆観光サービスにおいて、コミュニケーションロボットやARを活用している例もあります。

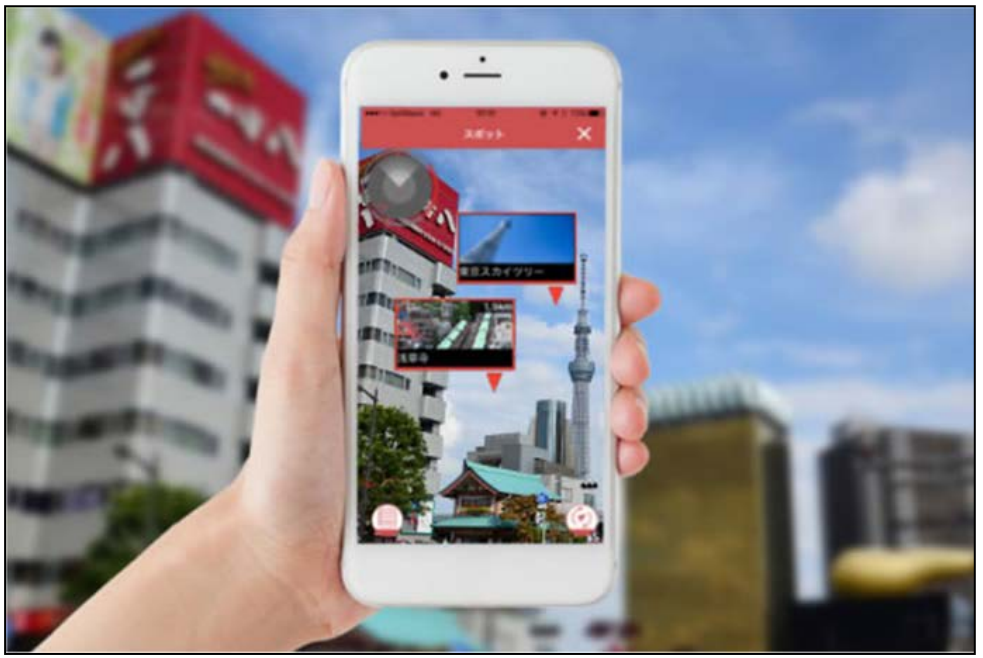
- 株式会社JTBは、株式会社シャープが販売しているコミュニケーションロボットRoBoHoN（ロボホン)を活用する京都の旅行ツアーを2017年12月～2018年3月に販売しました。
  - この企画は、経済産業省および一般社団法人日本ロボット工業会の「平成29年度ロボット導入実証事業」の一環として行われました。

RoBoHoNと行くロボ旅 京都のチラシ

【出所】モバイル型ロボット「RoBoHoN(ロボホン)」を使った京都の観光案内事業で協業［株式会社JTB西日本 ］
https://www.jtbcorp.jp/scripts_hd/image_view.asp?menu=news&id=00049&news_no=524

- 株式会社エム・ソフトが発売している観光ARアプリケーション「かんぷら」では、観光地の位置情報等に基づき、スマートフォン上にAR（拡張現実）を表示したり、プッシュ通知を行ったりできます。

「かんぷら」におけるARの表示例



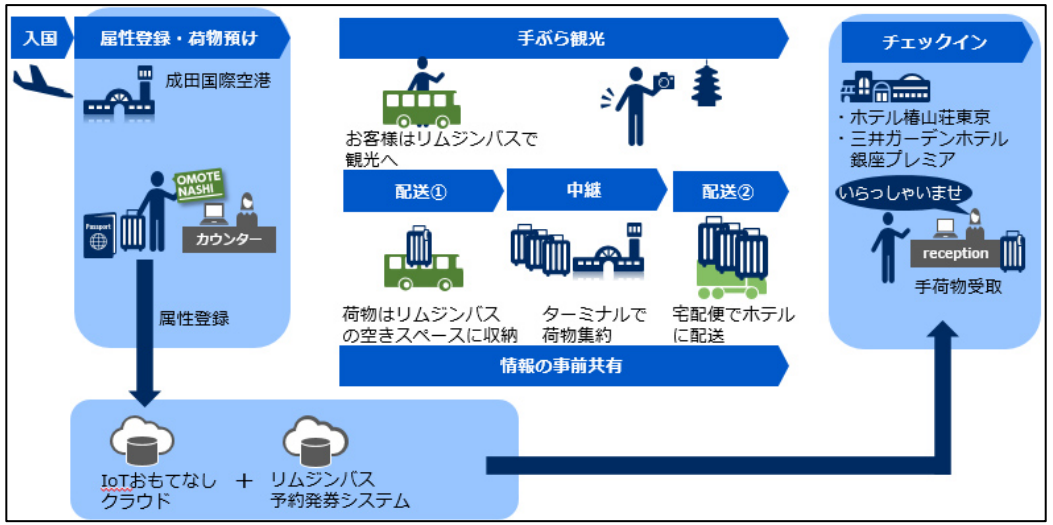
【出所】観光ARアプリケーション「かんぷら」［株式会社エム・ソフト］
https://kampla.biz-ar.jp/introduction/

(3)観光（訪日外国人向けサービス）

◆訪日外国人に焦点を当てた観光サービスにおいても、ICTが活用されています。

- 東京空港交通株式会社、ジャパンショッピングツーリズム協会、日本電気株式会社は、2018年2月に訪日外国人のリムジンバス受付時に「IoTおもてなしクラウド」を介して、ホテルと情報連携を行う実証実験を行いました。
  - 「IoTおもてなしクラウド」は、訪日外国人の利便性を向上に資するサービス連携を行うための共通クラウド基盤です。

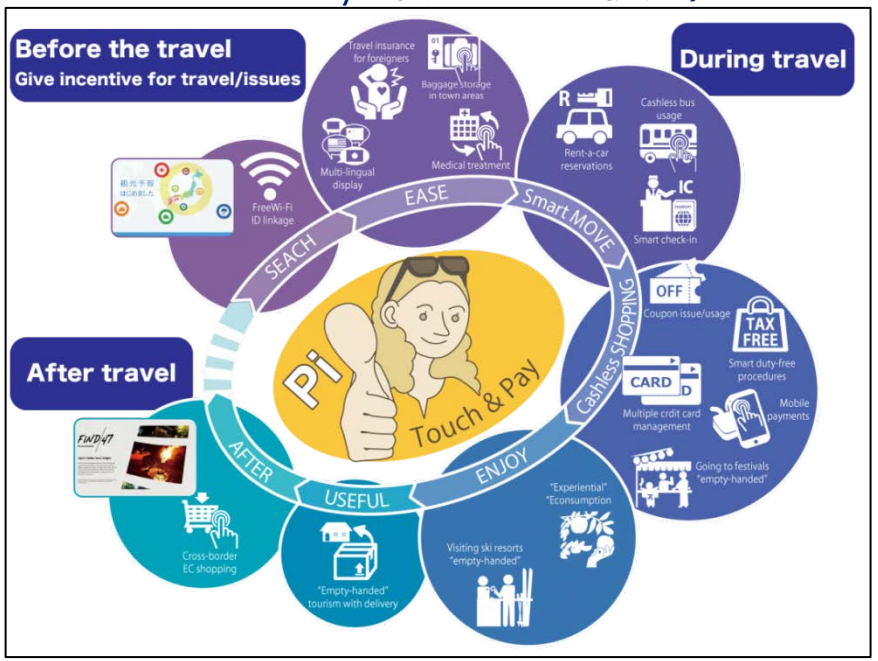
リムジンバスとホテルの情報連携の実証実験概要



【出所】訪日外国人に対するICTを活用したおもてなしサービスの実証実験を開始［一般社団法人ジャパンショッピングツーリズム協会］https://jsto.or.jp/wp-content/uploads/6d14a24a19534cc643d340f0b8dbe883.pdf

- Touch&Payは、訪日外国人向けの指紋認証システムで、神奈川県の湯河原温泉、兵庫県の城崎温泉、福岡市の博多等で利用することができます。
  - 「Touch&Pay」は、経済産業省が推進する「IoT活用おもてなし実証事業」の中で活用されているサービスの一つです。

Touch&Payの利用イメージ（英語）



【出所】about Touch&Pay［Touch&Pay事務局］https://touchandpay-japan.com/aboutus/

(4)金融（銀行のオープンAPIでの情報連携）

◆改正銀行法が公布され、銀行等のオープンAPIに関する義務が定められました。

- 2017年6月2日に銀行等の**オープンAPI**に関する義務を示した**改正銀行法が公布**され、銀行等には**オープンAPIに関する方針を公表する義務**、**オープンAPIに関する体制を整備する努力義務**が課されました。

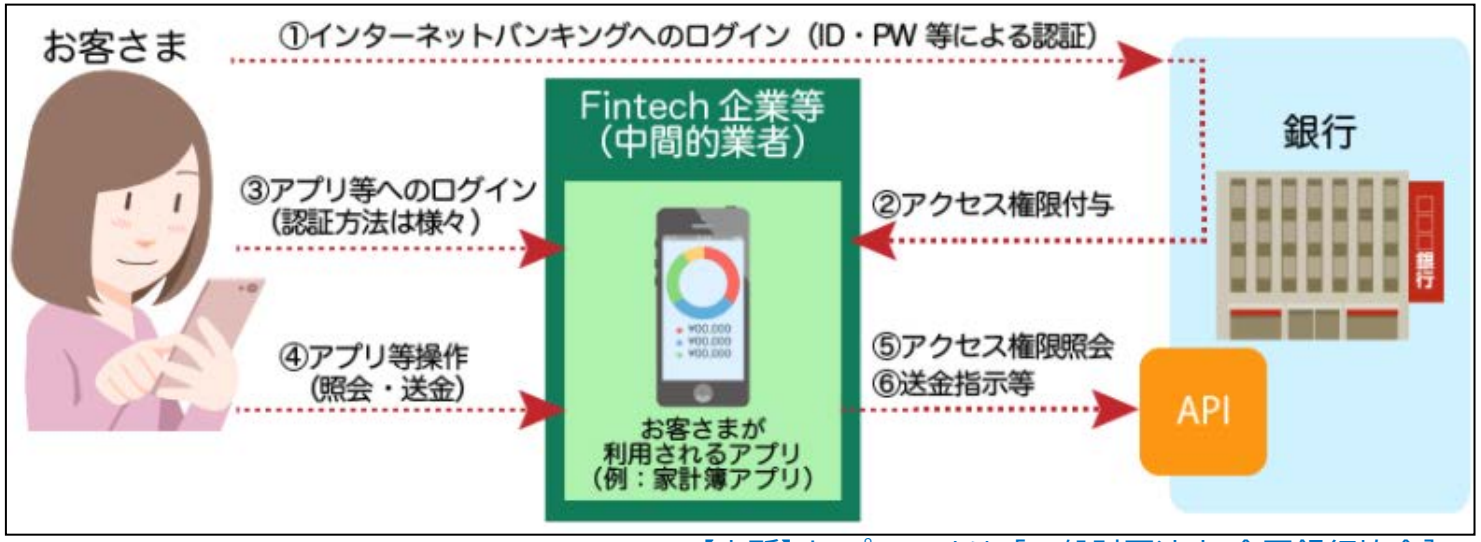
第10条　銀行等は、公布の日から起算して九月を経過する日までに、主務省令で定めるところにより、電子決済等代行業者等との連携及び協働に係る方針を決定し、これを公表しなければならない。

第11条 電子決済等代行業者等との間で（中略）契約を締結しようとする銀行等は（中略）当該銀行等に係る電子決済等代行業等を営むことができるよう、体制の整備に努めなければならない。

【出所】改正銀行法 附則（平成29年6月2日 法律第四九号）［e-Gov|総務省］
http://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/viewContents?lawId=356AC0000000059_20180601#AP

- 銀行等がオープンAPIの情報連携を開始すれば、**Fintech企業**（中間的業者）が新たなサービスを展開できます。

銀行におけるオープンAPIの基本的な仕組み



【出所】オープンAPIとは［一般財団法人 全国銀行協会］
https://www.zenginkyo.or.jp/article/tag-g/9797/

(4)金融（金融サービス等で利用されるブロックチェーン技術）

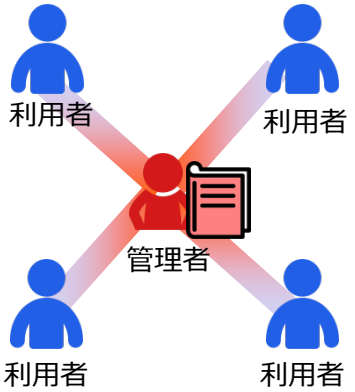
◆ブロックチェーン技術は、分散台帳を利用するデータの改ざんが困難な仕組みです。

- **ブロックチェーン**において、一定時間内における取引等の情報のやりとりを**ブロック**と呼びます。
- 各期間のブロックは、その**1期間前の暗号を含めて形成**されており、**ブロックの情報が鎖のようにつながっていることから、ブロックチェーン**と呼ばれます。
  - 新たなブロックは、proof-of-work（プルーフオブワーク）と呼ばれる時間のかかる計算作業とその作業に関わったコンピューターリソースの過半数の承認を経て、ブロックチェーンに追加されます。
  - 1期間前の暗号を含めたブロックが連なっているため、ある期のブロックの改ざんには、後続するブロックの暗号を全て変更する必要があります。
- ブロックチェーンの取引情報は、**ネットワークの参加者全員が閲覧できる分散台帳に保存**されています。
  - 従来のシステムにおいて主流であった中央集権型の管理では、管理者のみが全てのデータを保有し、各利用者とやりとりをする形でしたが、分散台帳においては、データは複数の利用者のコンピューターで重複して保存されています。
  - ブロックチェーン上に個人情報は記録されていないため、第三者はネットワークで共有された情報から取引を行った個人を特定できません。

ブロックチェーンにおけるブロックの繋がり方のイメージ

ブロックチェーンで利用される分散台帳

従来のシステムにおいて主流であった中央集権型の管理



□ ブロックチェーンのデータ構造やネットワーク上の承認体制から、データの改ざんが極めて困難な仕組みとなっています。

(4)金融（ブロックチェーン活用の推進団体と仮想通貨）

◆ブロックチェーンは幅広い注目を集めており、仮想通貨として決済に利用することができます。

● ブロックチェーンに関わる社団法人として「ブロックチェーン推進協会」「日本ブロックチェーン協会」が挙げられます。

- ブロックチェーン推進協会（BCCC）は2016年4月25日に設立され、ブロックチェーン技術の利用に関する情報交換・普及を主な目的としています。
- 日本ブロックチェーン協会（JBA）は2016年4月27日に設立され、仮想通貨サービスに関するガイドラインの策定や、ブロックチェーン技術の社会インフラへの応用、国へ政策提言を行い、産業発展に寄与することを主な目的としています。

● **ブロックチェーンの代表的な利用例**として、ビットコインなどの**仮想通貨**が挙げられます。

- 2008年11月に著者Satoshi Nakamotoより、公表されたビットコインに関する論文において、ブロックチェーンの仕組みが初めて示されました。

【出所】Bitcoin: A Peer-to-Peer Electronic Cash System［Satoshi Nakamoto］ https://bitcoin.org/bitcoin.pdf

● ビットコインは、ビックカメラなど店舗やネットショッピングサイトにおける**決済に利用できるケース**があります。

- ビックカメラの店舗のビットコインの支払いは、店舗が所有するタブレット等に表示されたQRコードを読み取り、送金を行います。

ブロックチェーン推進協会のロゴ

Blockchain Collaborative Consortium

【出所】日本ブロックチェーン協会
http://bccc.global/ja/

日本ブロックチェーン協会のロゴ

JBA

【出所】日本ブロックチェーン協会
http://jba-web.jp/

ビックカメラにおける
ビットコインの決済に関する告知

【出所】株式会社ビックカメラ
http://www.biccamera.co.jp/shopguide/campaign/bitcoin/

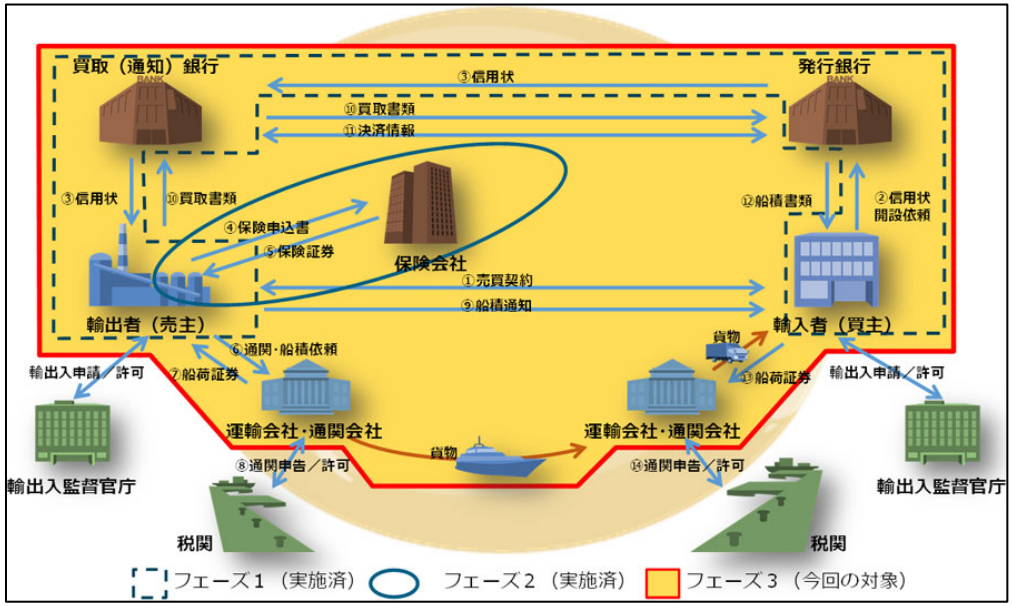
【参考】ブロックチェーン技術の広範な活用

◆ブロックチェーン技術は金融サービス以外にも様々な応用が期待されています。

- ブロックチェーン技術は**「データの改ざんが極めて困難な仕組みとなっていること」**や**「分散台帳によって、データ消失やサービス停止の懸念が極めて小さいこと」**から、**金融サービス以外でも活用**されようとしています。

- 株式会社NTTデータが事務局を務める「ブロックチェーン技術を活用した貿易情報連携基盤実現に向けたコンソーシアム」では、貿易に関する手続き・情報連携をブロックチェーン技術によって効率化するための実証を行っています。

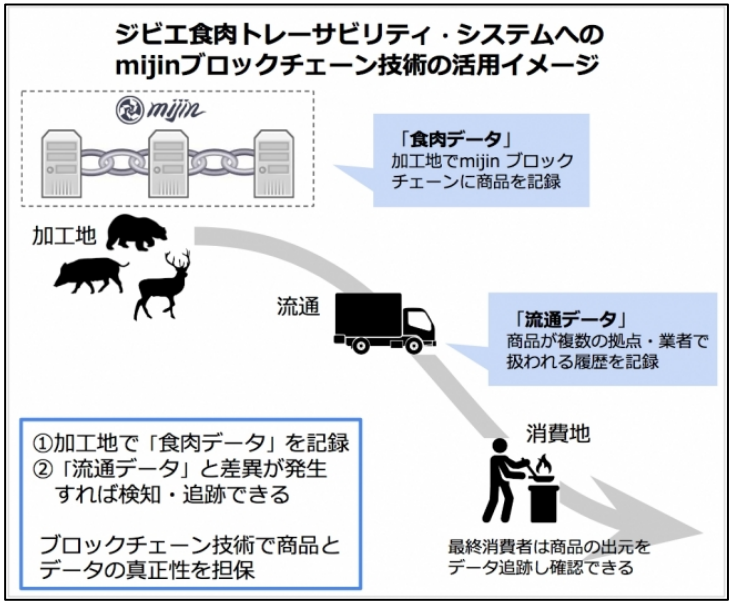
ブロックチェーンを利用する貿易情報連携基盤の実証



【出所】ブロックチェーン技術を活用した貿易情報連携基盤の実現に向け、13社でコンソーシアムを発足株式会社［株式会社NTTデータ］
http://www.nttdata.com/jp/ja/news/services_info/2017/2017081501.html

- 一般社団法人日本ジビエ振興協会とテックビューロ株式会社は、ブロックチェーン技術に基づくジビエ食肉のトレーサビリティシステムを構築し、2017年10月より試験運用を行っています。

ブロックチェーンを利用する食肉トレーサビリティ



【出所】テックビューロ株式会社
http://mijin.io/ja/985.html

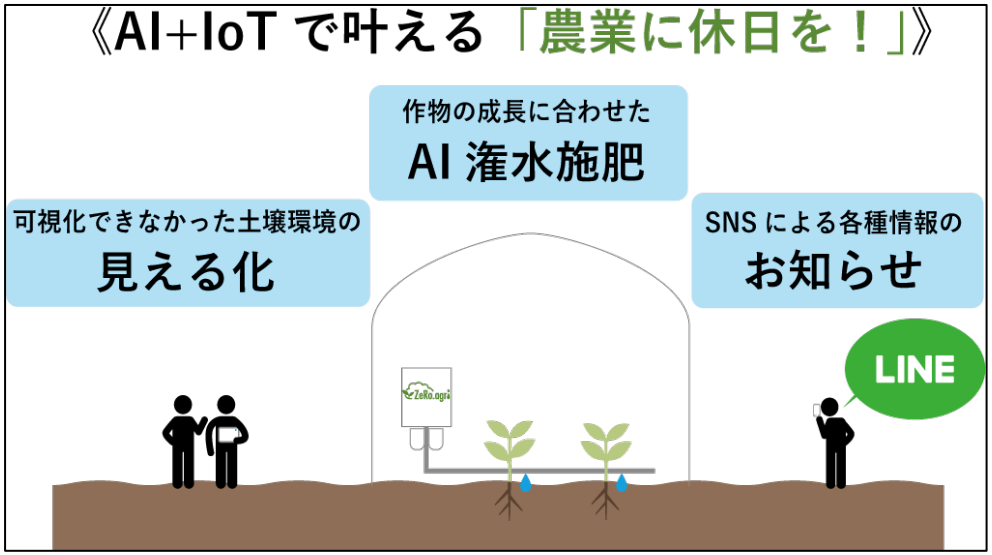
(5)農林水産（農業におけるIoTの活用）

◆農業においても、IoTやデータ分析の活用が進展しています。

- 株式会社ルートレック・ネットワークスが提供する次世代養液土耕システム「ゼロアグリ」では、各種センサー情報からAIが現在の作物状況にあった最適な液肥供給量を判断し、自律的に供給します。
  - 熟練農業者がゼロアグリの液肥供給量を調整することで、栽培技術を後継者にスムーズに伝えることも可能となります。

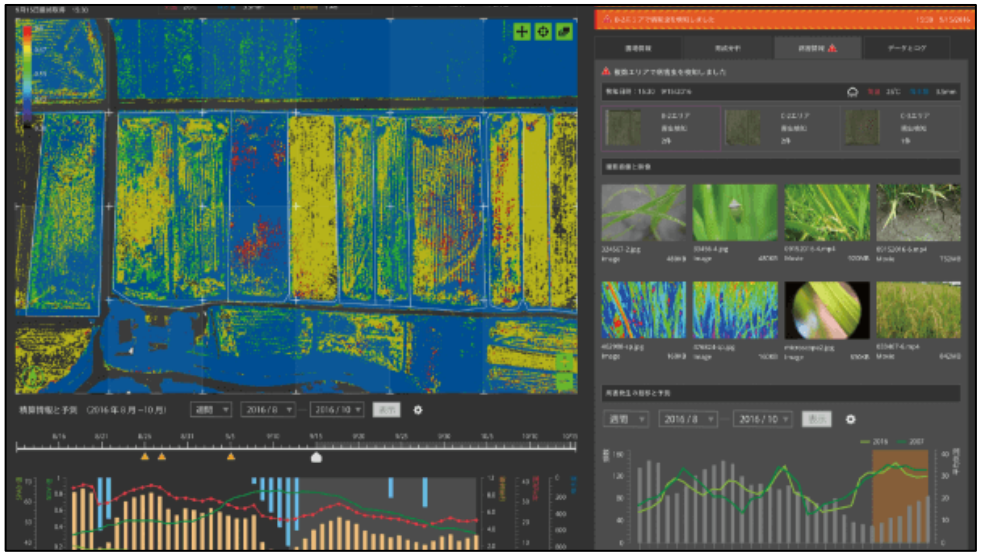
- 株式会社オプティムが提供している圃場情報管理サービス Agri Field Manager（アグリ フィールド マネージャー）では、カメラ等のセンサーから得られたビッグデータを分析し、作物の生育の可視化や予測を行います。
  - 画像データをディープラーニングで処理することで、作物の数をカウントしたり、病害を検知することも可能です。

ゼロアグリの基本機能



【出所】ゼロアグリとは［株式会社ルートレック・ネットワークス］
http://www.zero-agri.jp/about/

Agri Field Managerによる画像・センサーデータの分析



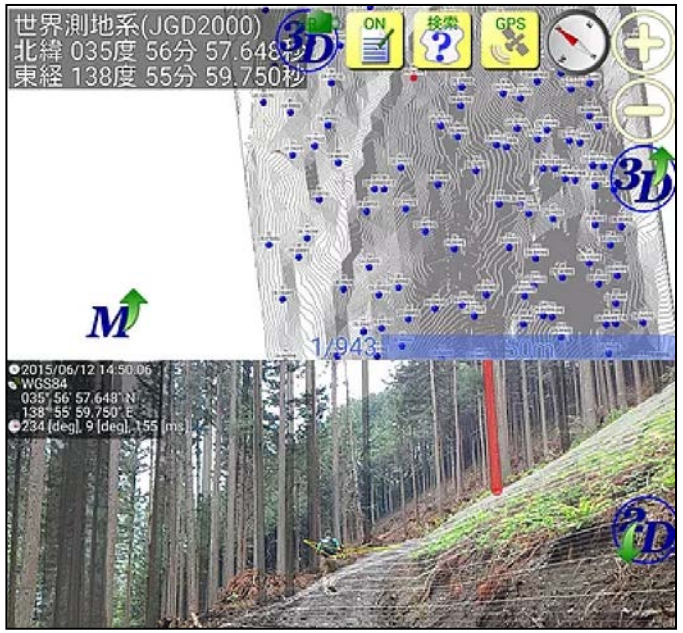
【出所】圃場情報管理サービス Agri Field Manager［株式会社オプティム］
https://www.optim.co.jp/agriculture/agri-field.php

(5)農林水産（林業・水産業におけるIoTの活用）

◆林業・水産業では、IoTを活用して、持続的経営を可能にする情報を収集しています。

- 株式会社woodinfoのシステムDigital Forestでは、林内の全木の位置と太さ、高さ、体積、曲り等を自動で情報化します。
- 各木の性質・特徴を把握することで、伐採する立木の計画を行い、AR（拡張現実）を使って、目指す立木まで案内します。

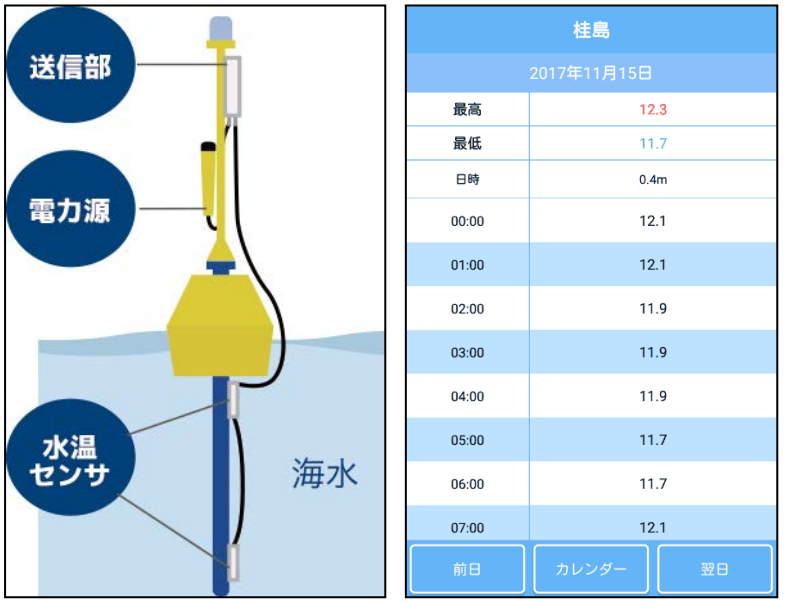
スマートフォンによる立木へのナビ



【出所】木（もく）ナビシステム［株式会社woodinfo］
http://maple96.wixsite.com/woodinfo/services2-cixd

- アンデックス株式会社は、公立はこだて未来大学の和田雅昭教授と協力し、ユビキタスブイによる海中の水温等の計測とスマートフォンアプリによる情報発信を行っています。
  - 収集した各地の水温情報はユビキタスブイシステム（JellyFish）のウェブサイトから、無料で閲覧することができます。

ユビキタスブイとスマートフォンにおける水温の表示



【出所】ユビキタスブイとは［アンデックス株式会社］
http://marine-it.net/activity/ubiquitous-buoy.shtml

(6)ものづくり（工場におけるウェアラブルデバイスの活用）

◆工場の製造過程においても、ウェアラブルデバイスは活用されています。

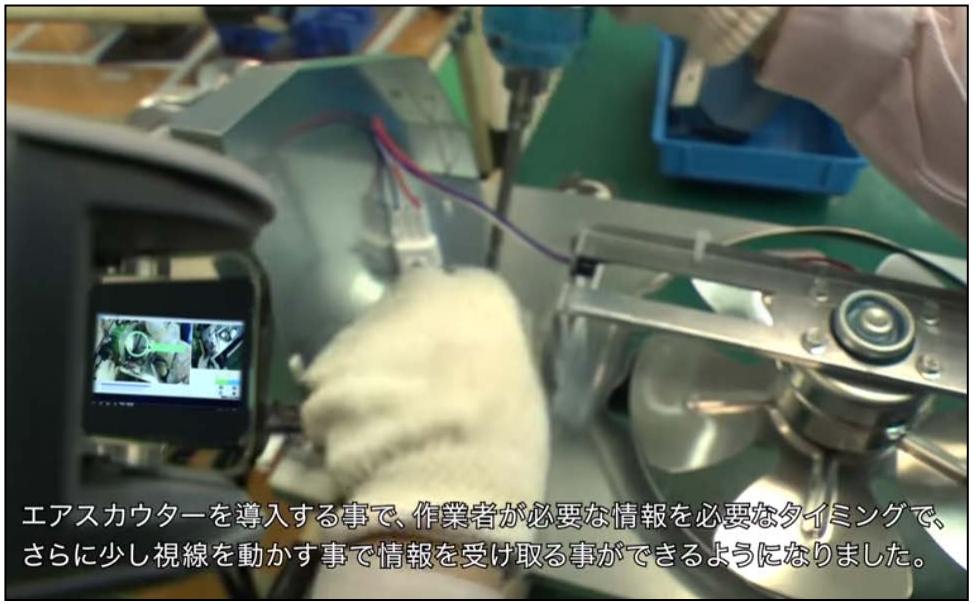
- ブラザー工業株式会社が発売しているエアスカウターは、ウェアラブルデバイスとして頭に装着することで、視線の脇にあるディスプレイとして利用することができます。
  - 利用者はドローンを操縦しながら、ドローンのカメラから送信された画像をエアスカウターで見ることもできます。

エアスカウター（WD-200A）

【出所】WD-200A［ブラザー工業株式会社］
http://www.brother.co.jp/product/hmd/wd200a/index.aspx

- 冷蔵ショーケースなどを生産しているパナソニック株式会社の群馬大泉工場では、工場の作業者にエアスカウター着用させました。作業に必要な情報を必要なタイミングで得られるようにすることで、作業者の生産性の向上を図りました。

パナソニックの工場におけるエアスカウターの利用



【出所】AiRScouter(エアスカウター) 導入事例：パナソニック群馬大泉工場［ブラザー工業株式会社］
https://www.youtube.com/watch?v=oEiH-lb7nnM

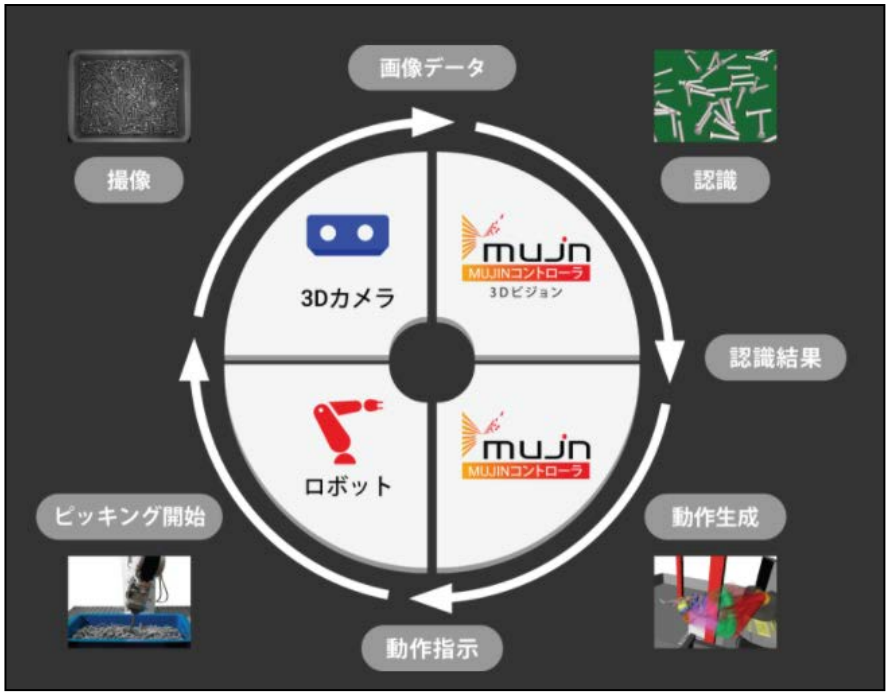
(6)ものづくり（産業用ロボットにおけるAI活用）

◆AIを搭載した産業用ロボットは、臨機応変な稼働・対応が可能となります。

- 株式会社MUJINが提供しているロボットコントローラーでは、3Dカメラと人工知能によって、様々な形状の対象物に対して、プログラム不要で臨機応変な対応が可能となります。

- MUJINによってコントロールされた産業用ロボットは、事前の登録作業は不要で、大きさや形状が異なる不定形の箱を次々にピッキングして、移動させることができます。

MUJINによる画像認識と自律動作の流れ



【出所】株式会社MUJIN
https://mujin.co.jp/

不定形の箱をピッキングし、移動させる産業用ロボット

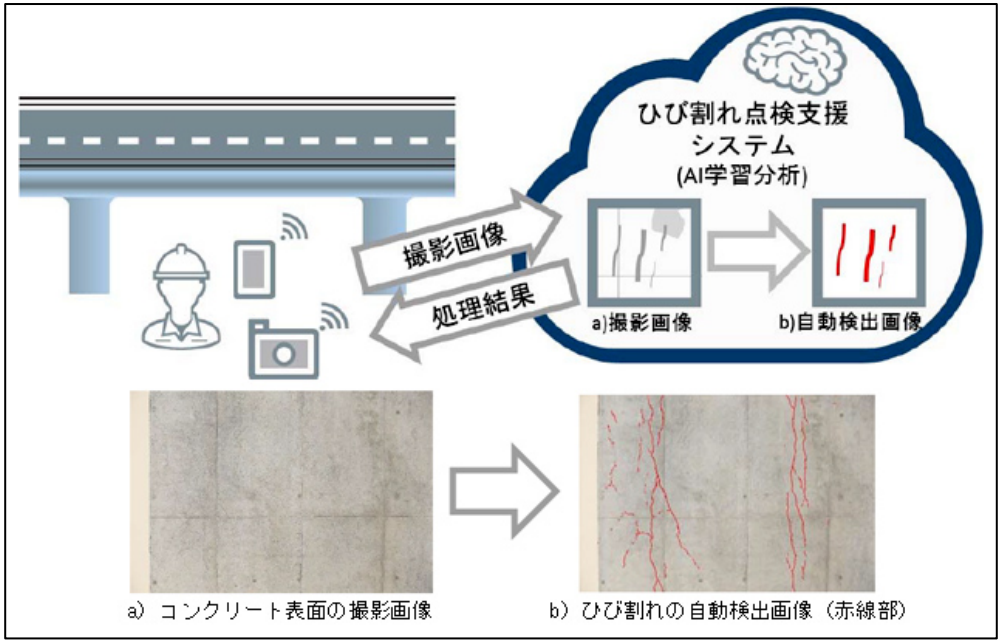
【出所】株式会社MUJIN
https://www.youtube.com/watch?v=A0vObFJOS64

(7)インフラ・防災・減災（インフラストラクチャー部門のAI活用）

◆インフラストラクチャー部門の保守や安全管理へのAI活用の実証実験も行われています。

- 国立研究開発法人新エネルギー・産業技術総合開発機構のプロジェクトでは、コンクリートひび割れを高精度で検出するAIを開発し、有効性を検証しています。
- 2017年8月から2018年度末（予定）まで「ひび割れ検出Webサービス」として無料公開されています。

AIに基づくコンクリートのひび割れ点検支援システム



【出所】国立研究開発法人新エネルギー・産業技術総合開発機構(NEDO)
http://www.nedo.go.jp/news/press/AA5_100811.html

- 株式会社オプティムは、JR九州の一部の駅において、「AI Physical Security Service（エーアイフィジカルセキュリティサービス）」の実証実験を2017年10月より開始し、駅構内の異常検知、安全性の向上に関する有効性を検証しています。

駅に設置したカメラ画像の解析イメージ

【出所】株式会社オプティム
https://www.optim.co.jp/news-detail/23884

(7)インフラ・防災・減災（VRを利用した防災・減災）

◆災害や避難訓練をVR（拡張現実）で体験することで、防災・減災に役立ちます。

- 森トラスト株式会社では2018年3月の震災訓練の一環として、**VRを通じた災害体験、避難訓練**を実施しました。
- 白山工業株式会社が開発している体験型地震シミュレーター「地震ザブトン×VR」を利用すれば、過去に発生した大震災や将来想定される震災の揺れなど6種類を体験することが可能です。
- 「地震発生後に火災が発生し、煙が立ち込める中、誘導灯を頼りに火災現場から脱出すること」をVRによって体験することができました。

「地震ザブトン×VR」の体験の様子

【出所】森トラスト株式会社
https://www.mori-trust.co.jp/pressrelease/2018/20180313.pdf

VRによる火災時の疑似避難訓練の様子

【出所】森トラスト株式会社
https://www.mori-trust.co.jp/pressrelease/2018/20180313.pdf

(8)移動（コネクティッドカー）

◆自動車がIoTとなることによって、段階的に自動運転が可能となっていきます。

- **インターネットにつながっている自動車は「コネクティッドカー」**と呼ばれ、走行、メンテナンス、渋滞に関する情報の送受信を行うことができます。
- コネクティッドカーは、インターネットから受信した情報と「センサーが感知した情報」「人工知能による制御・運転」と組み合わせることで、**運転支援や高速道路における自動運転が、段階的に可能**になっていきます。
- 首相官邸が公開している**「官民 ITS 構想・ロードマップ 2017」**では、自動運転の段階を「運転支援」～「完全自動運転」までの5段階に分けて達成時期を示しています。
  - ITSは、「Intelligent Transport Systems」の頭文字に由来しており、日本語では「高度道路交通システムと訳されます。
  - 官民 ITS 構想・ロードマップ 2017が示した自動運転の段階の基準は、米国道路交通安全局（NHTSA）が2016年9月に発表したSAE（Society of Automotive Engineers）に準拠しています。
  - SAEでは、自動車の運転自動化のない「レベル0」から一般道を含めた完全自動運転を達成する「レベル5」があり、「レベル0」を除外すると、5段階の自動運転に関する段階があります。

官民 ITS 構想・ロードマップ2017における自動運転システムの市場化・サービス実現の期待時期

	レベル	実現が見込まれる技術（例）	市場化等の期待時期
自家用車	SAE レベル 2（部分運転自動化）	準自動パイロット	2020年まで
	SAE レベル 3（条件付運転自動化）	自動パイロット	2020年目途
	SAE レベル 4（高度運転自動化）	高速道路における完全自動運転	2025年目途
物流	SAE レベル 2（部分運転自動化）以上	高速道路におけるトラックの隊列走行	2022年以降
	SAE レベル 4（高度運転自動化）	高速道路におけるトラックの完全自動運転	2025年以降

【出所】「官民 ITS 構想・ロードマップ 2017（自動運転システムの市場化・サービス実現期待時期）」に基づき作成
https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20170530/roadmap.pdf

(8)移動（運転手の眠気検知センサー・IoT自転車）

◆IoTはバス運転手の眠気の検知やレンタルバイクとしても、活用されています。

- WILLER EXPRESS JAPAN株式会社では、高速バスの運転手が、疲れや眠気を検知するウェアラブルセンサーを利用することで、2017年における事故による修繕費は、前年比74%減になりました。
  - 富士通株式会社が販売するウェアラブルセンサーFEELythm（フィーリズム）は、運転手本人への警告だけでなく、運行管理者への通知も可能です。

高速バスの運転手が利用するウェアラブルセンサー

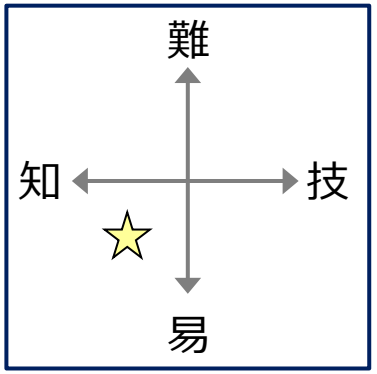
【出所】株式会社富士通マーケティング
https://www.youtube.com/watch?v=o2fz6Bym8X0

- 株式会社ドコモ・バイクシェアでは、GPSを搭載したIoT自転車を用いたレンタルサイクル事業を展開しており、利用者は30分単位などで、自転車を借りることができます。
  - 株式会社ドコモ・バイクシェアが運営するバイクシェアにおいて、利用者は各駐輪場にある利用可能な自転車の台数をスマートフォンから確認することができます。

バイクシェアに利用されるIoT自転車

【出所】株式会社ドコモ・バイクシェア
http://www.d-bikeshare.com/

総務省　ICTスキル総合習得教材



［コース4］オープンデータ・ビッグデータ利活用事例

4-5：データ利活用の発表・交流の場

http://www.soumu.go.jp/ict_skill/pdf/ict_skill_4_5.pdf

	1	2	3	4	5
［コース1］データ収集					
［コース2］データ蓄積					
［コース3］データ分析					
［コース4］データ利活用					●

本講座の学習内容［4-5：データ利活用の発表・交流の場］

【講座概要】

- 「官民データ活用推進基本法」における人材育成、普及啓発の方針を紹介します。
- ICT人材の育成政策、ICT活用の普及啓発の政策を示します。
- アイデアソン、ハッカソンなどのICT、データの利活用に関する学習・交流の場を紹介します。
- RESASの政策コンテストなどのICT、データの利活用に関する発表・競争の場を紹介します。
- 地方版IoT推進ラボなどのICT、データの利活用に関する実践・実証の場を紹介します。

【講座構成】

座学

- [1] 日本政府のICT人材育成政策
- [2] ICT・データ利活用に関する学習・交流の場
- [3] ICT・データ利活用に関する発表・競争の場
- [4] ICT・データ利活用に関する実践・実証の場

【学習のゴール】

- ✓ 「官民データ活用推進基本法」における人材育成、普及啓発の方針を理解する。
- ✓ ICT人材の育成政策、ICT活用の普及啓発の政策を把握する。
- ✓ ICT、データの利活用に関する「学習・交流の場」「発表・競争の場」「実践・実証の場」の事例をそれぞれ紹介できる。

日本におけるICT人材の育成・普及啓発

◆日本政府はICT・データを活用できる人材育成・普及啓発に責務を持っています。

- 2016年12月に施行された**「官民データ活用推進基本法」**では、日本政府が官民データ活用に関する**「人材の育成及び確保」「教育及び学習の振興、普及啓発」**に必要な措置を講じることが規定されています。
  - 「官民データ活用推進基本法」と「世界最先端ＩＴ国家創造宣言・官民データ活用推進基本計画」は、講座4-4において紹介しました。

> **第17条**　国は、官民データ活用に係る専門的な知識又は技術を有する人材を育成し、及び確保するために必要な措置を講ずるものとする。
>
> **第18条**　国は、国民が広く官民データ活用に関する関心と理解を深めるよう、官民データ活用に関する教育及び学習の振興、啓発及び知識の普及その他の必要な措置を講ずるものとする。

【出所】官民データ活用推進基本法［e-Gov|総務省］
http://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/detail?lawId=428AC1000000103

- 2017年5月に閣議決定された**「世界最先端ＩＴ国家創造宣言・官民データ活用推進基本計画」**においても、人材育成に関して、次のように示されています。

> **人材育成、普及啓発等**
>
> - AI、IoT 等を有効に活用するために不可欠なデータ活用に係る専門的な知識や技術を有する人材の育成について、政府一体となって計画的に実施
> - 不足するセキュリティ・IT 人材の計画的な育成
> - 社会人の学び直しの推進（技術系人材の再教育）

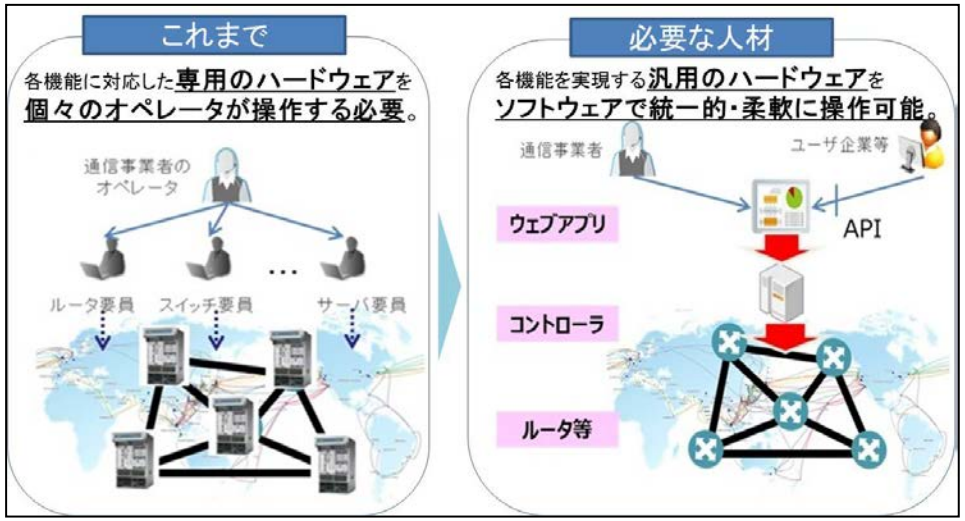
【出所】世界最先端IT国家創造宣言・官民データ活用推進基本計画［高度情報通信ネットワーク社会推進戦略本部（ＩＴ総合戦略本部）］
http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20170530/siryou1.pdf

ICT人材の育成に関する政策方針

◆IoT活用やサイバーセキュリティ確保に関する人材育成の政策方針が示されています。

- 膨大なIoT機器等が迅速・効率的にネットワークに接続するために必要な技術を運用する**人材を育成する環境基盤を整備**します。
- 環境基盤の構築・運用を通して人材育成を図り、求められる**スキルの明確化**や**その認定の在り方**を検討します。
  - 技術の進展・普及によってネットワークの運用・管理に求められる人材・技能も変化しています。

IoT時代のネットワークを運用・管理に求められる人材の変化



- 巧妙化・複合化するサイバー攻撃に対し、実践的な対処能力を持つセキュリティ人材を育成するため、国立研究開発法人情報通信研究機構(NICT)を通じて、以下の取組を実施します。

①国の行政機関、地方公共団体、独立行政法人及び重要インフラ事業者等に対するサイバー攻撃について、実践的なサイバー防御演習(CYDER)

②2020年東京オリンピック・パラリンピック競技の適切な運営に向けたセキュリティ人材の育成(サイバーコロッセオ)

③若手セキュリティイノベーターの育成(SecHack365)

新たなサイバー攻撃に対応するための取り組み

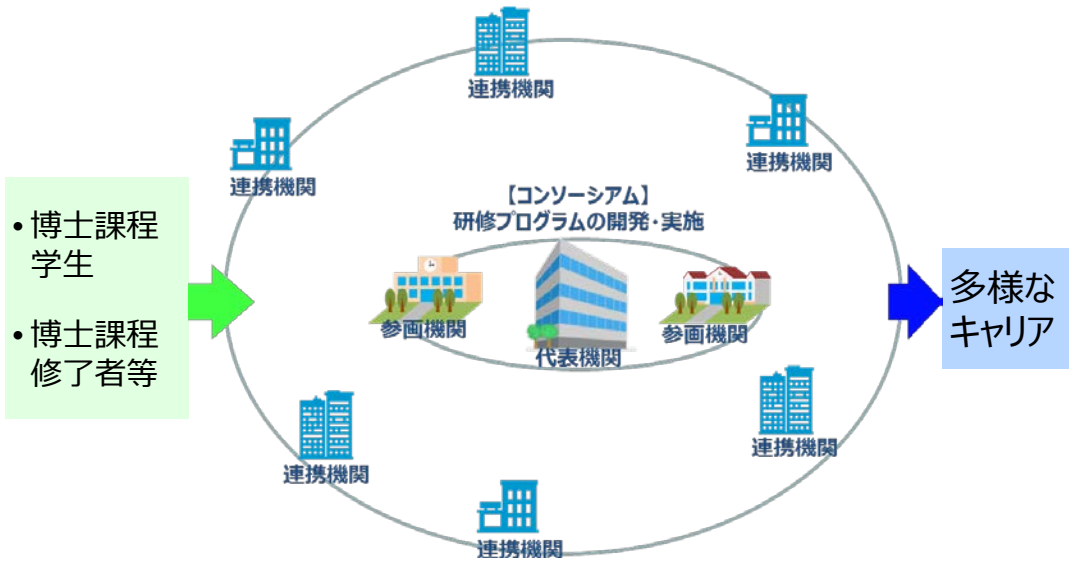
【出所】落ち着いて、やさしく、持続可能な社会の実現（総務省重点施策2018） – 重点施策集 –
［総務省］http://www.soumu.go.jp/main_content/000506557.pdf

ICT人材の育成に関する政策

◆日本政府は様々な人材に対して、ICTに関する技能形成を支援しています。

- 文部科学省では2017年度から**「データ関連人材育成プログラム」**として、博士課程学生や博士号取得者等を対象とするデータ関連技術（AI、IoT、ビッグデータ、セキュリティ等）のキャリア開発プログラムを実施しています。
- 「データ関連人材育成プログラム」では、企業や大学等がコンソーシアムを形成し、人材の育成やキャリア開発支援を行っています。
- 2018年4月から認定講座が開始する**「第四次産業革命スキル習得講座（Reスキル講座）」**は、経済産業大臣が情報通信技術・データ活用を習得できる講座を認定する制度です。
- 厚生労働省の**「教育訓練給付制度」**と連携しており、厚生労働省の指定も受けている講座は受講費用の一部が支給される場合もあります。

「データ関連人材育成プログラム」の実施イメージ



【出所】データ関連人材育成プログラム［文部科学省］
http://www.mext.go.jp/a_menu/jinzai/data/

「第四次産業革命スキル習得講座」の事業者懇談会の様子

【写真提供】経済産業省
http://www.meti.go.jp/policy/economy/jinzai/reskillprograms/

ICT活用の普及啓発に関する政策

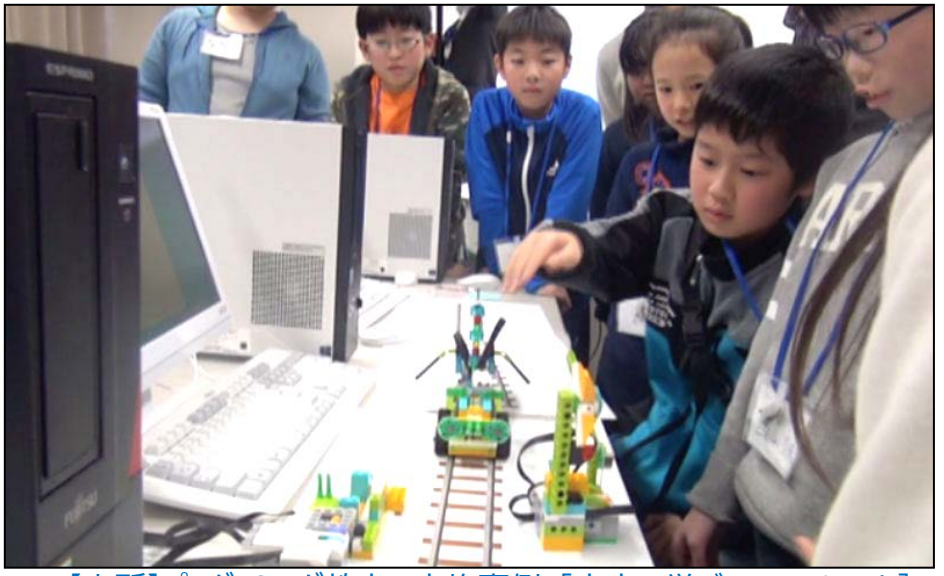
◆日本政府は、プログラミング教育を中心にICT人材の普及啓発政策を推進しています。

- 文部科学省では**「次世代の教育情報化推進事業」**として、情報活用能力の育成に関するカリキュラム・マネジメントの在り方等の調査研究を実施しています。
- 「次世代の教育情報化推進事業」の中では、教員に対してプログラミング教材の研修を行うなど、効果的なICT教育の実施支援・検討をすることもあります。
- 文部科学省、総務省、経済産業省が連携し、2017年3月にプログラミング教育の普及・促進のため**「未来の学びコンソーシアム」**を設立しました。
- 「未来の学びコンソーシアム」は、ポータルサイトが設置され、その中では総務省の「若年層に対するプログラミング教育の普及推進」事業などが紹介されています。

事業説明会における教員へのプログラミング教材の研修の様子

【出所】「次世代の教育情報化推進事業」説明会［つくば市総合教育研究所］
http://www.tsukuba.ed.jp/~souken/?p=7226

「若年層に対するプログラミング教育の普及推進」事業の様子



【出所】プログラミング教育の実施事例［未来の学びコンソーシアム］
https://miraino-manabi.jp/search/

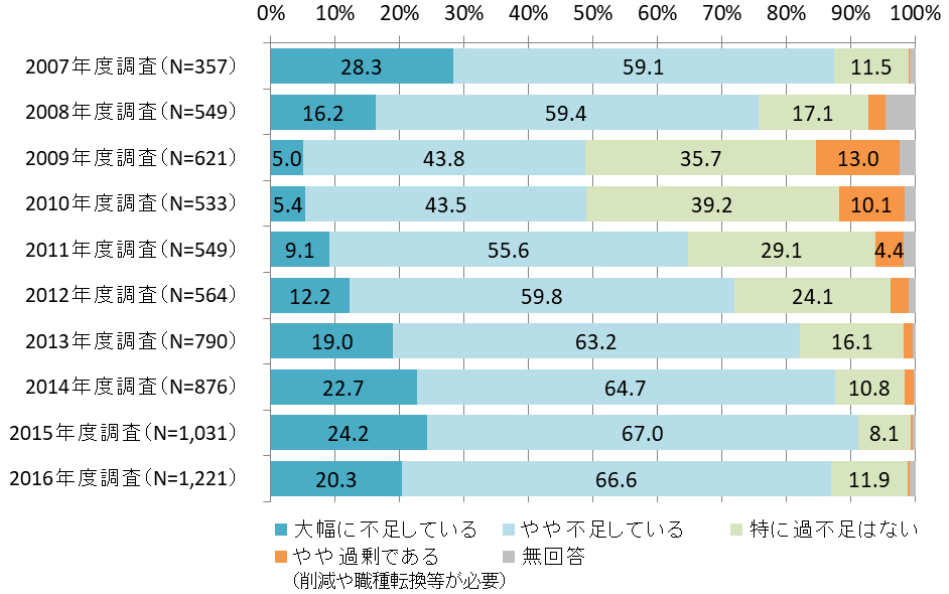
□ ICT人材の育成、普及啓発政策が実施されていますが、ICT人材が不十分であることを示すデータもあります。

日本の企業におけるIT人材の不足感

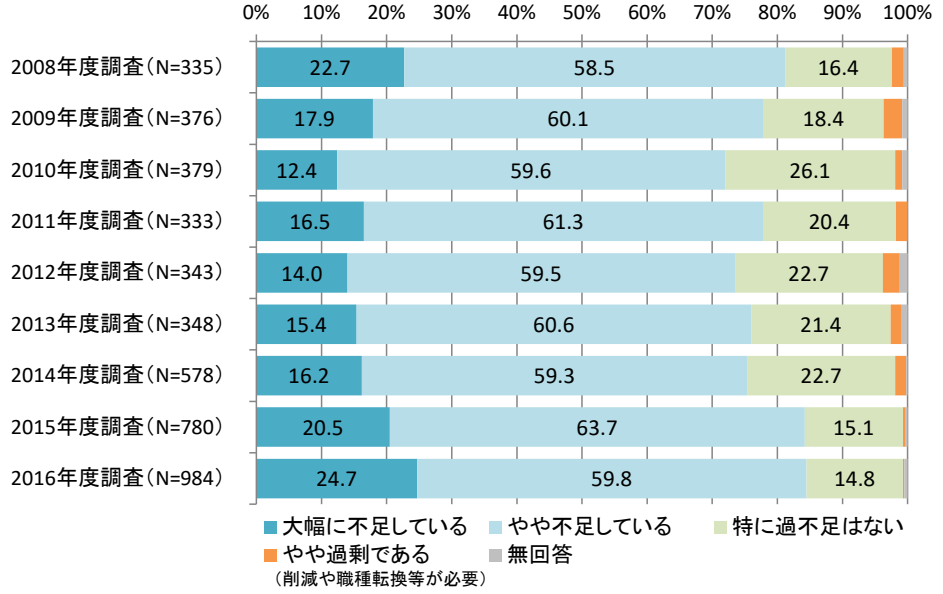
◆日本の企業におけるIT人材の量に対する不足感は、増加傾向にあります。

- （独）情報処理推進機構では、「IT人材動向調査」として、「IT企業」と「ユーザー企業」にIT人材に関する過不足感のアンケート調査を実施しています。
  - 本教材では主にICT（情報通信技術）という言葉を使っていますが、ICTとIT（情報技術）は、概ね同じ意味で使われています。
  - 「IT人材動向調査」では、ITの提供側の企業を「IT企業」、ITの利用側の企業を「ユーザー企業」と呼び、企業データベースの業種（情報・通信）等から分類しています。「IT人材動向調査」の詳細な結果は、無料で公表されているIT人材白書に記載されています。
- 日本におけるIT人材の不足感は、IT企業、ユーザー企業ともに過去5年で増加傾向にあり、2016年調査において、不足を感じている割合は8割を超えています。
  - 不足を感じている割合は「大幅に不足している」「やや不足している」の合計と見なせば、2016年調査においてIT企業では86.9%、ユーザー企業では84.5%です。

IT企業のIT人材の“量”に対する過不足感（10年推移）



ユーザー企業のIT人材の“量”に対する過不足感（9年推移）

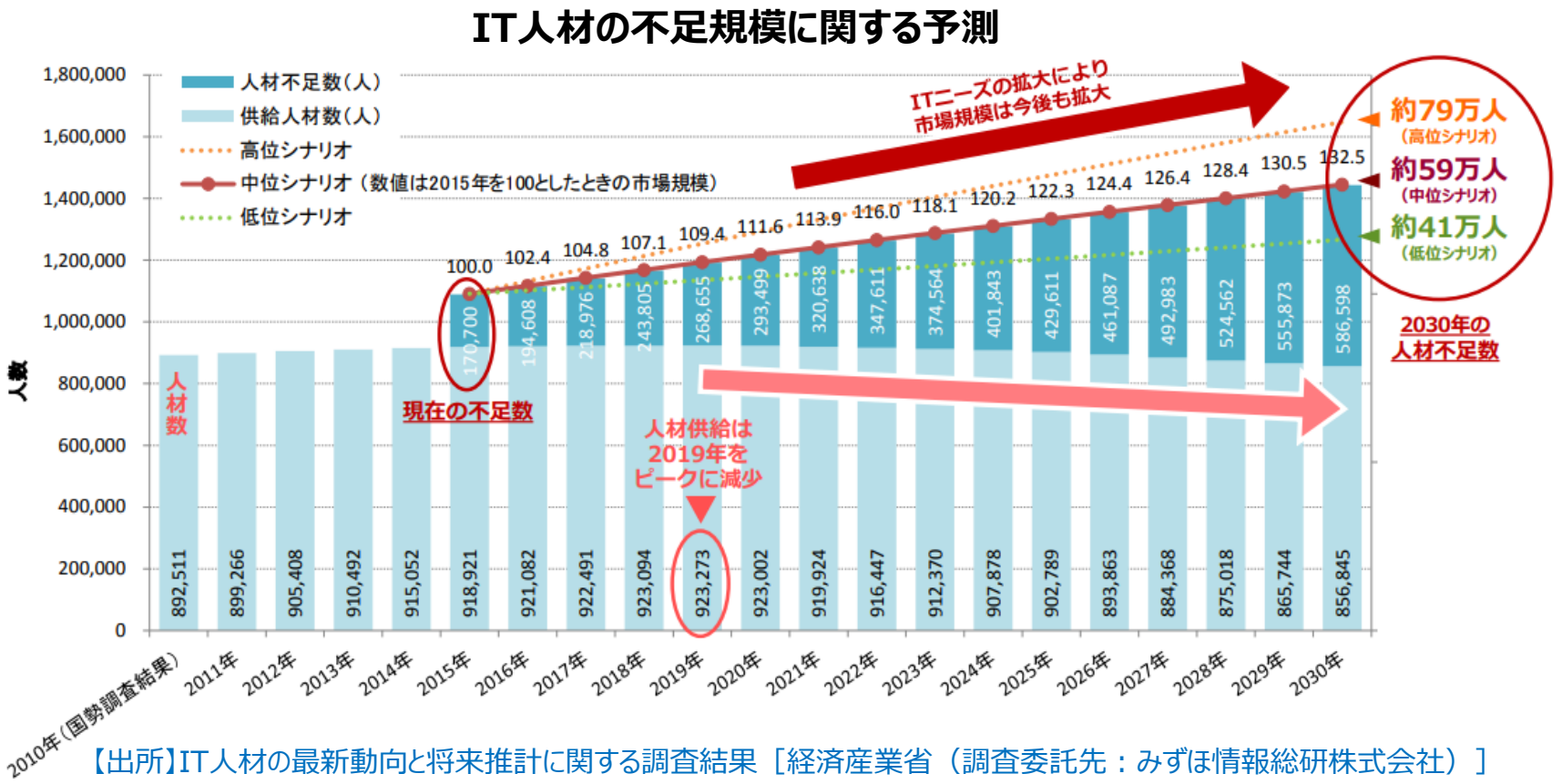


【出所】IT人材白書2017［（独）情報処理推進機構］の公表データに基づき作成 https://www.ipa.go.jp/jinzai/jigyou/about.html

IT人材の不足規模に関する予測値

◆IT人材は現在も不足し、今後不足規模が増加する予測がされています。

- 経済産業省が委託した調査の結果によれば、2015年時点のIT人材の不足規模は17万人でしたが、2030年の不足規模（中位シナリオ）は約59万人になるとの予測がされています。



【出所】IT人材の最新動向と将来推計に関する調査結果［経済産業省（調査委託先：みずほ情報総研株式会社）］
http://www.meti.go.jp/policy/it_policy/jinzai/27FY/ITjinzai_report_summary.pdf

- IT人材の不足規模の拡大が見込まれる中、様々な取り組みを通して知識や技能を高める必要があります。

地域IoT官民ネット

◆地域IoT官民ネットは、地域におけるIoT実装を支援する団体です。

- 2017年7月に自治体、民間企業、総務省が連携し、地域におけるIoT実装を推進するために**「地域IoT官民ネット」**が設立されました。
  - 地域におけるIoT実装に関心を持つ自治体、民間企業は無料で参加でき、2018年3月時点で119の自治体、181の民間企業が加盟しています。
- 地域IoT官民ネットのウェブサイトでは、これまでに実施した勉強会等の様子が示されています。
- 地域IoT官民ネットでは、IoTやデザインシンキングに関心がある女性自治体職員、女性民間事業者社員を「IoTデザインガール」として募り、女性によるIoT活用や活用支援を推進しています。

「地域IoT官民ネット」のウェブサイト

【出所】地域IoT官民ネット
http://www.local-iot.jp/

「第四次産業革命スキル習得講座」の事業者懇談会の様子

【出所】地域IoT官民ネット
http://www.local-iot.jp/

ハッカソン、アイデアソン

◆ハッカソンやアイデアソンは、より良いアプリ、アイデアを生み出すイベントです。

- **ハッカソン**や**アイデアソン**は、**ハック(hack)**、**アイデア(idea)** と**マラソン(marathon)**を組み合わせた混成語です。
  - ハッカソンはプログラミングやデザインの経験者が集まって、テーマに沿ったアプリ制作を競う集まりです。
  - Yahoo! JAPANでは、1日目の正午から2日目の正午までの24時間を開発時間とする「Hack Day」を開催しています。
  - アイデアソンは、与えられた技術を活用するサービスや地域活性化のアイデアを議論する集まりで、一般にICTの専門知識は不要です。
    - アイデアソンで有益なアイデアが生まれた場合、アプリ制作の知識・技術を持つ者の制作を依頼するケースもあります。

Hack Day 2017の集合写真

【出所】Hack Day 2017［ヤフー株式会社］
https://hackday.jp/2017/

国土交通省が開催した
位置情報活用サービスを考えるアイデアソン

【出所】国交省が位置情報活用サービスを考えるアイデアソン開催［内閣府］
http://qzss.go.jp/news/archive/mlit_ideathon_161118.html

ハッカソン、アイデアソンの事例

◆日本各地で様々なハッカソン、アイデアソンが開催されています。

- 開催地やテーマがユニークなハッカソン、アイデアソンの事例を紹介します。

- SPAJAMは「温泉でハッカソン」を合言葉として、温泉地における2泊3日のスケジュールで、スマートフォンアプリを開発するハッカソンです。
  - SPAJAMは、一般社団法人モバイル・コンテンツ・フォーラムが主催し、地方予選を経てから、温泉地での本戦が開催されます。

- ロボットの社会実装を目指したアイデアを18歳以上の学生が検討する「福岡ロボットアイデアソン」が2018年1月に福岡市において開催されました。
  - 「福岡ロボットアイデアソン」は複数の企業・大学によって実行委員会が組織され、福岡市および北九州市における事前説明会を経て、開催されました。

箱根町で開催されたSPAJAM2017の様子

【出所】SPAJAM　https://spajam.jp/history/

福岡市で開催されたロボットアイデアソンの様子

【出所】株式会社システムトランジスタ　http://systrabot.com/?p=3072

ICTに関する勉強会・イベントが検索できるウェブサービス

◆ICTの勉強会、イベント等の開催情報が検索できるウェブサービスもあります。

- 日本の各地において、ICT活用やデータ分析に関する集まりが開かれています。
  - 情報検索サービスの利用は無料で行えますが、掲載されているイベントには参加料が必要なケースもあります。

- パーソルキャリア株式会社が運営する「TECH PLAY」では、IT勉強会やセミナーの情報を検索することができます。
  - 参加可能性のある都道府県や関心がある情報通信技術を事前に登録しておくことで、対象を絞った検索ができます。

TECH PLAY

【出所】TECH PLAY［パーソルキャリア株式会社］
https://techplay.jp/

- 株式会社ビープラウドが運営する「connpass」は、IT勉強会支援プラットフォームとして、申込時の参加料の決済サービスなど、イベント運営者を支援する機能があります。

connpass

【出所】connpass［株式会社ビープラウド］
https://connpass.com/

ICTに関する勉強会・イベントが検索できるウェブサービスの事例

◆勉強会・イベント検索サービスでは、地域や技術を絞った検索が可能です。

- TECH PLAYにおいて「沖縄県」に絞った検索を行えば、那覇市における「初心者歓迎」のイベントが見つかりました。
  - 地方においても、様々な勉強会・イベントが開催されていることに加えて、検索サービスを活用して、勉強会・イベント情報を発信することが可能です。
  - 「初心者歓迎」の勉強会、作業や雑談が中心となっているイベントを見つけることもできます。

那覇市における「初心者歓迎」のイベント告知（TECH PLAY）

【出所】TECH PLAY［パーソルキャリア株式会社］
https://techplay.jp/

- connpassにおいて「RESAS」で検索すると、「RESASとR言語で学ぶ統計データ入門」の勉強会が見つかりました。
  - 小規模な勉強会では、特定の技術やプログラミング言語に特化することが容易で、専門的な内容を掘り下げる学習や意見交換に向いています。

「RESASとR言語で学ぶ統計データ入門」のイベント告知（connpass）

【出所】connpass［株式会社ビープラウド］
https://connpass.com/

コンテスト・フォーラム

◆コンテスト、フォーラムでは、成功事例を聞いたり、アイデアを話すことができます。

- **「地方創生 政策アイデアコンテスト」**ではRESASの分析結果に基づいて、地域振興のアイデアのプレゼンテーションを行います。
- 「地方創生 政策アイデアコンテスト2017」では、小学生が最優秀賞に相当する地方創生担当大臣賞（高校生以下の部）を受賞しました。
  - 例年、「高校生以下の部」と「大学生以上の部」の2部門で応募者を募っています。

「地方創生 政策アイデアコンテスト2017」にて
地方創生担当大臣賞を受賞した小学生

【出所】内閣府 地方創生推進室
https://contest.resas-portal.go.jp/2017/

- 内閣府 地方創生推進室では、RESASに関するフォーラムやイベントを日本各地で開催し、RESASを通じた意見交換を促進しています。
  - RESASについては講座4-2にて紹介しましたが、地域情報の分析ツールのみならず、地域活性化に関するコミュニケーションツールとしても活用されています。

「地方創生 RESASフォーラム 2017 in 関西」の告知

【出所】内閣府 地方創生推進室
https://forum.resas-portal.go.jp/2017kansai/

情報通信技術に関する展覧会

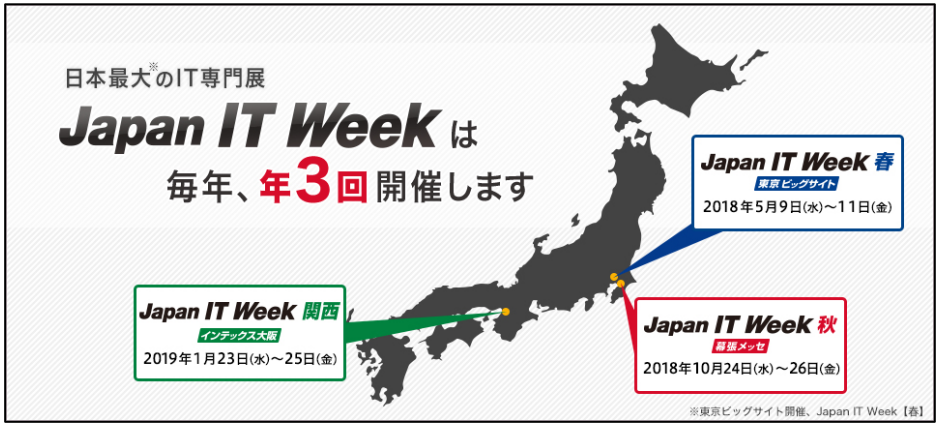
◆情報通信技術に関する展覧会にて、新製品・サービスの告知、協業の相談が可能です。

- 民間企業が主催する情報通信技術に関する展覧会においては、新製品・新サービスを出展、観覧することにより情報共有ができ、協業のための商談が可能です。

- リード エグジビション ジャパン株式会社が主催する「Japan IT Week」は、複数の専門展覧会で構成され、年3回開催されています。

- 株式会社日経BPが主催するIT専門展示会のイベントは、2018年において東京都、大阪府、福岡県、愛知県、北海道の全国5箇所で開催予定となっています。

2018年における「Japan IT Week」の告知と構成



2018年におけるJapan IT Week春を構成する専門展覧会		ソフトウェア&アプリ開発展
組込みシステム開発技術展	情報セキュリティEXPO	データセンター展
クラウド コンピューティングEXPO	通販ソリューション展	AI・業務自動化展
ビッグデータ活用展	データストレージEXPO	Web&デジタルマーケティングEXPO
モバイル活用展	IoT/M2M展	店舗ITソリューション展

【出所】Japan IT Week［リード エグジビション ジャパン株式会社］
http://www.japan-it.jp/

日経BP主催のIT展示会の構成と出展対象（2018年春～夏）

出展対象製品・サービス

Cloud Days 2018
- クラウド導入/構築/運用サービス
- クラウド向けアプリケーションサービス（SaaS）
- クラウド基盤サービス（PaaS、IaaSなど）
- クラウドAIサービス
- サーバ、ストレージ、ネットワーク機器など
- 仮想化ソフトウエア、仮想化環境構築サービス
- 運用管理、データセンターなど

Security 2018
- 脅威対策ソリューション（ウィルス/スパイウエア対策など）
- 情報漏洩対策ソリューション（暗号化ツール、ログ管理など）
- アクセス制御ソリューション（ワンタイムパスワード、認証デバイスなど）
- 物理セキュリティソリューション（入退出管理、監視画像システムなど）
- クラウドセキュリティ、仮想環境セキュリティ、データセンターセキュリティ・セキュリティコンサルティングサービス、セキュリティ監査サービス

IoT Japan 2018
- IoTシステム/プラットフォーム
- IoTアプリケーション/サービス
- IoT関連ソリューション（施設管理、防災/セキュリティ、販売/物流、車載/運輸、医療/ヘルスケア、建設/土木などIoT/M2Mセンサネットワーク技術）
- 無線通信技術、無線モジュール
- ビッグデータ関連（データの管理・分析・活用・生成・蓄積・統合・保護・保管・破棄など）
- セキュリティシステム関連製品
- 制御・計測システム関連製品
- スマートファクトリー関連製品
- スマートオフィス/ホーム関連製品
- ウェアラブル関連製品・技術
- コンサルティング、導入支援サービス

FACTORY 2018
※東京・関西・名古屋会場のみ
- 出展対象製品・サービス
- 製造業向け情報システム
- 製造業向けデータ関連ソリューション
- 製造業向けセキュリティ
- FA／PA／生産システム
- 設計／開発／計測・評価
- デバイス／モジュール
- サプライ・チェーン
- コンサルティング／アウトソーシング

ビジネスAI 2018
- ディープラーニング
- 機械学習
- ニューラルネットワーク
- 自然言語処理
- 画像認識
- 音声・音響認識
- ビッグデータ関連
- ロボット関連
- マーケティング関連
- コンテンツ制作関連
- AI応用ハードウエア/システム
- AI応用アプリケーション/サービス
- RPA(ロボティック・プロセス・オートメーション)
- RDA(ロボティック・デスクトップ・オートメーション)

その他、各種バックオフィス・業務自動化関連ソリューション

働き方改革 2018
- 在宅勤務/テレワーク関連ソリューション
- モバイルワーク関連ソリューション
- ペーパーレス関連（デジタル複合機、ペーパーレス会議システム など）
- 会議システム関連関連ソリューション
- コミュニケーション関連（電子メール、グループウエア、ビジネスチャットツール、スケジュール管理ツール など）
- 業務サポート関連（意思決定支援、営業支援、名刺管理、経費/出張精算、電子承認/決済 など）
- フリーアドレスオフィス関連ソリューション

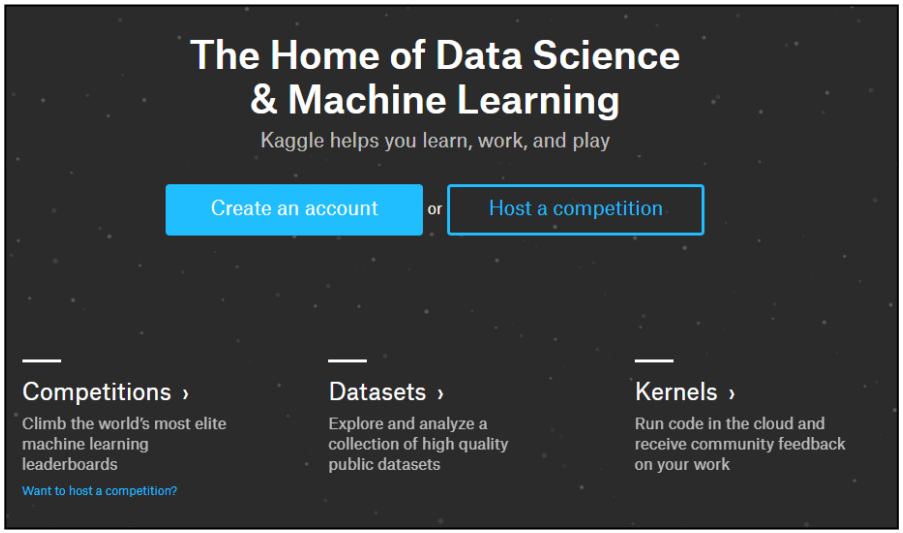
【出所】Nikkei BP　ITイベント［株式会社日経BP］
http://expo.nikkeibp.co.jp/cloud/2018/exhibition/

データ分析力を競うウェブサイト

◆データ分析の技術や精度を競うことができるウェブサイトがあります。

- Kaggle（カグル）は世界規模でデータ分析の技術を競うためのウェブサイトです。
  - Googleは2017年3月にKaggleを買収しましたが、Googleのクラウドの中でKaggleのブランド名は継続されています。
- 企業や団体が競争を主催し、成績優秀者に賞金を出すケースもあれば、データサイエンスの腕試しのための課題が用意されるケースもあります。
  - 機械学習の分析結果の精度を競い合い、分析に関する成績によって世界規模でのランキングが公表されています。

Kaggle



【出所】Kaggle ［Kaggle Inc］
https://www.kaggle.com/

- SIGNATE（シグネイト）では、企業・行政は課題と懸賞金を設定し、データを匿名化して ウェブサイトに設置して、分析結果を募ります。
- データサイエンティストはSIGNATEに掲載されたデータを分析し、コンテスト形式で分析モデルの精度を競い合います。

SIGNATE

【出所】SIGNATE［株式会社SIGNATE］
https://signate.jp/

Kaggle、SIGNATEにおける分析課題例

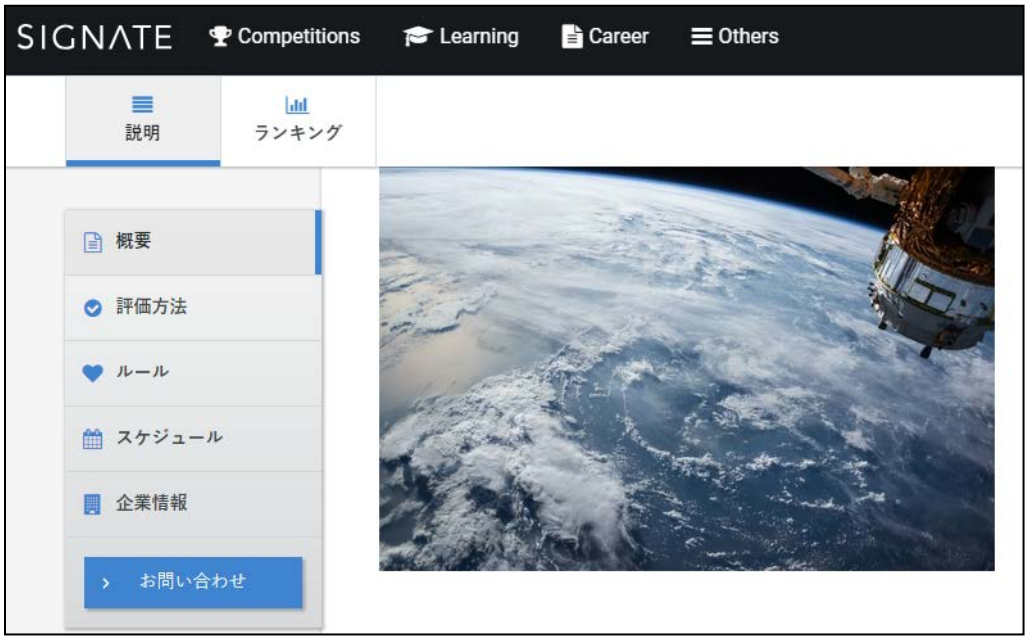
◆賞金が設定されている分析課題は、ビジネスの課題や社会課題の解決に貢献するものです。

- フリーマーケットアプリを運営する株式会社メルカリは、2017年11月に出品時の価格査定の精度改善に関する分析課題をKaggleにて公開しました。
  - この分析課題の賞金として、1位には＄60,000、2位には＄30,000、3位には＄10,000が設定されました。
- SIGNATEのコンテストとして、国立研究開発法人の産業技術総合研究所が設定した「衛星画像からゴルフ場を検出」する事例が挙げられます。
- 実際の社会課題・ビジネスの課題の解決に貢献するテーマに賞金が設定され、分析を公募しています。

Kaggleにおける株式会社メルカリの賞金付き課題

【出所】MERCARI PRIZE［Kaggle Inc］
https://kaggle.mercari.com/2017/

SIGNATEにおけるコンテスト（画像からゴルフ場の検出）



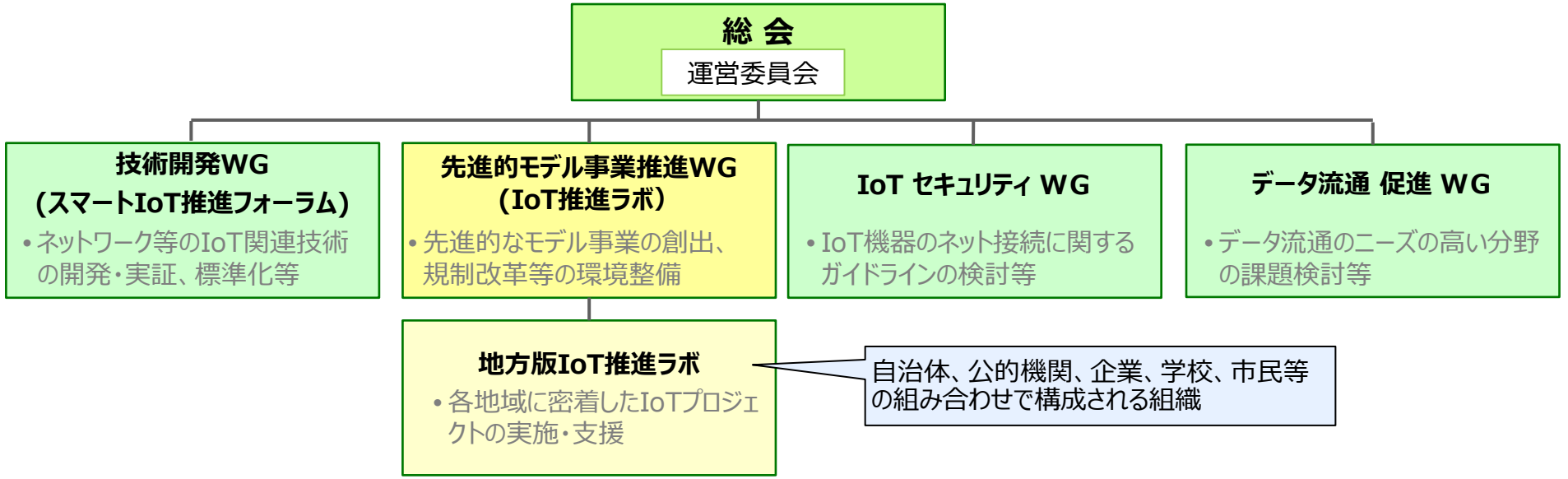
【出所】産業技術総合研究所 衛星画像分析コンテスト
［株式会社SIGNATE］https://signate.jp/competitions/61

IoTに関する実践・実証支援をするIoT推進ラボ

◆IoTコンソーシアムの下にはIoTの実践・実証を支援するIoT推進ラボがあります。

- **IoT推進コンソーシアム**の下で企業連携・資金・規制改革を支援するグループとして、**IoT推進ラボと呼ばれる先進的モデル事業推進ワーキンググループ**があります。
  - 講座1-1で示したように、IoT推進コンソーシアムは、産官学が一体となってIoT関連事業を実施、支援している組織です。
- 各地域に密着したIoTプロジェクトの実施・支援を行うべく、経済産業省とIoT推進ラボの連名で自治体等が母体となる**地方版IoT推進ラボ**を選定しています。
  - 「地方版IoT推進ラボ」は、自治体、公的機関、企業、学校、市民等の組み合わせで構成される組織であり、組織の事務局の自治体が応募資格を持つ形で公募されます。
  - 地方版IoT推進ラボは、自らビジネス化を目指す組織であるケース、ビジネス化を支援する組織であるケースがそれぞれあり得ます。

IoT推進コンソーシアムの組織構造（2017年9月時点）



【出所】2017年度（第3回）総会 の公開資料［IoT推進コンソーシアム］に基づき作成　http://www.iotac.jp/event/plenary2017/

IoT推進ラボの取り組み

◆IoT推進ラボは、先進的IoTプロジェクトの実践や企業連携を支援しています。

- IoT推進ラボは先進的なIoTの取り組みを表彰したり、IoTを活用するビジネスのマッチング支援を行ったりしています。

- IoT推進ラボが実施する**IoT Lab Selection**では、先進的なIoTプロジェクトを募集し、優秀な取り組みの表彰や支援を行っています。
  - 「成長性・先導性」「波及性（オープン性）」「社会性」「実現可能性」といった基準で審査が行われます。

第4回 IoT Lab Selectionの表彰式の様子

【出所】CEATEC JAPAN 2017におけるIoT推進ラボ［IoT推進ラボ］
https://iotlab.jp/jp/article/171006202531.html

- IoT推進ラボが実施する**IoT Lab Connection**では、IoT推進ラボの会員企業同士のビジネスマッチングや企業連携を支援しています。
  - IoT Lab Connection（Solution Matching ）では「観光」「製造」「ヘルスケア」といったテーマを設定し、企業連携を促しています。

IoT Lab Connectionにおけるビジネスマッチング

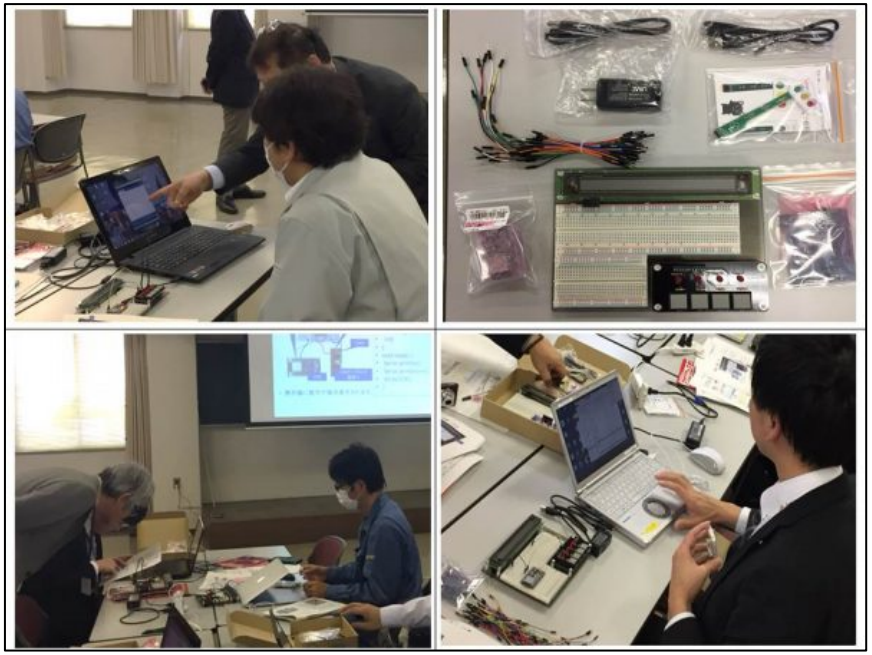
【出所】IoT Lab Connection［IoT推進ラボ］
https://iotlab.jp/jp/connection.html

地方版IoT推進ラボの取り組み

◆地方版IoT推進ラボは、各地域におけるIoTの利活用や人材育成を支援しています。

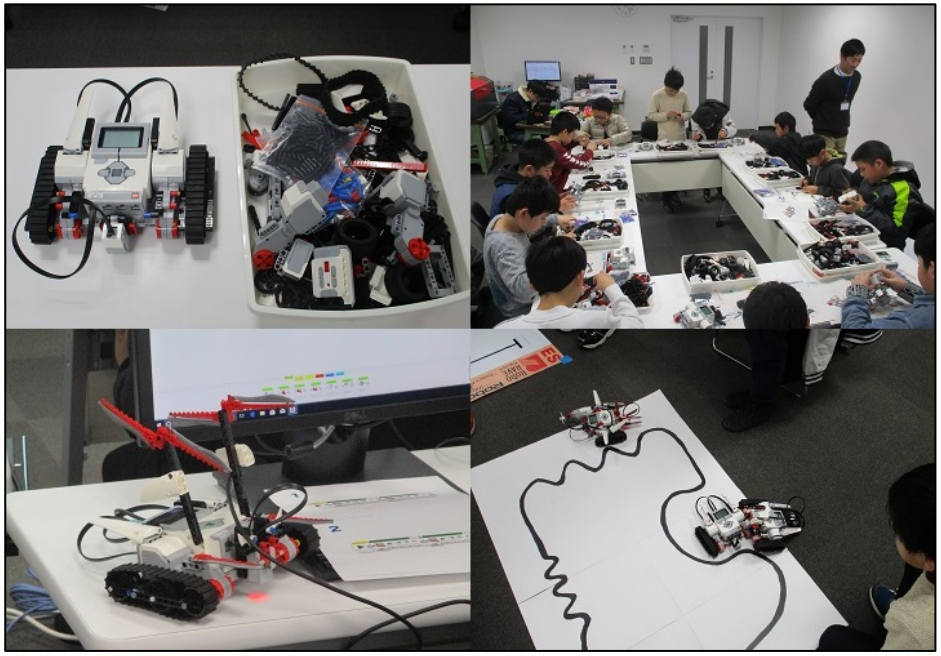
- 地方版IoT推進ラボでは、地域に密着したICT活用の講座を実施したり、普及啓発活動を行っています。
  - 三重県IoT推進ラボが、2018年3月に開催したIoTワークショップは、中小企業の従業員等を対象とし、IoT機器に温度センサーを取り付け室温をウェブサイトに表示させる実習などを行いました。
  - 石川県にある加賀市IoT推進ラボが、2018年3月に開催したIoTものづくり講座は、小学生を対象とし、車型ロボットの作成やロボットを走らせるプログラミングの実習を行いました。

三重県津市におけるIoTワークショップ



【出所】IoTワークショップ［三重県IoT推進ラボ］
https://local-iot-lab.ipa.go.jp/article/mie-pref-iot-10.html

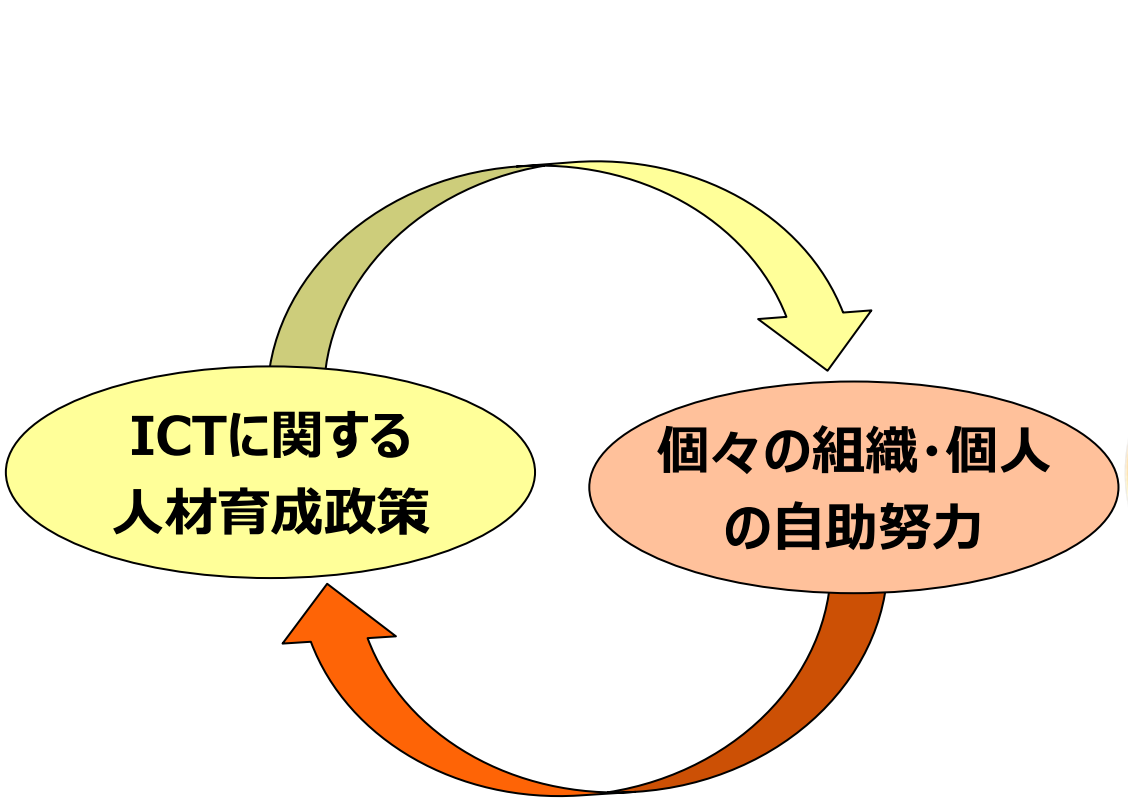
石川県加賀市におけるIoTものづくり講座

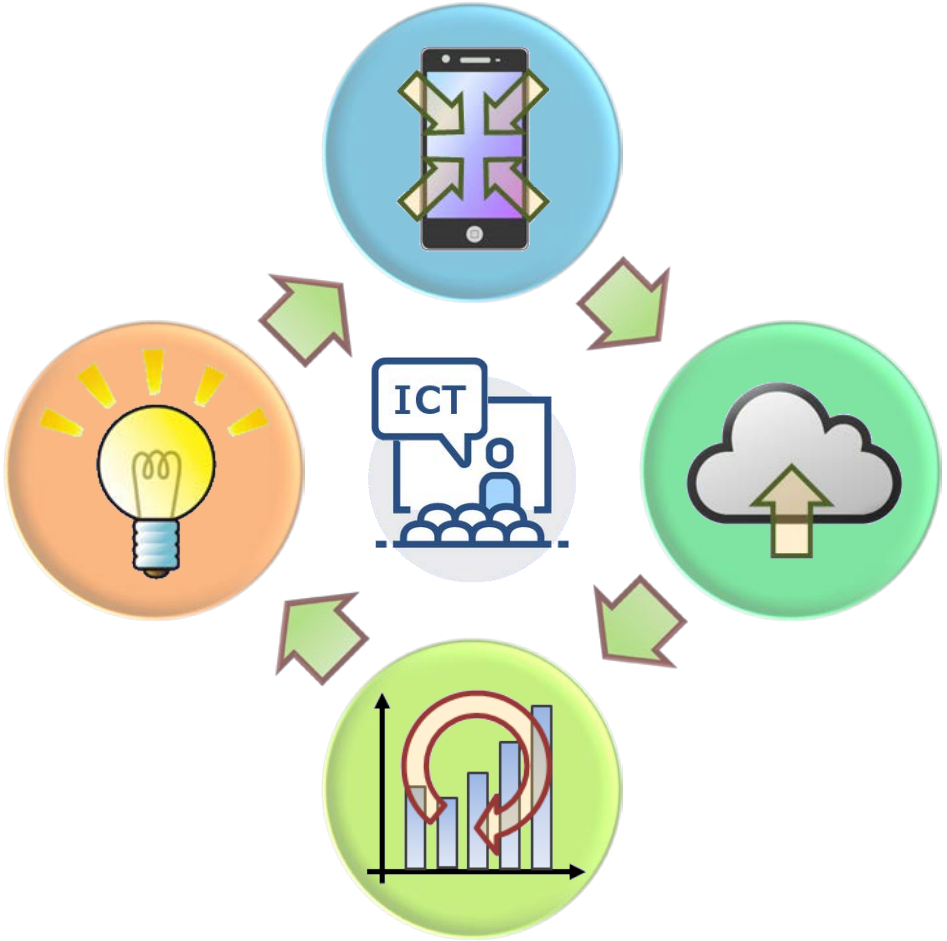


【出所】IoTものづくり講座［加賀市IoT推進ラボ］
https://local-iot-lab.ipa.go.jp/article/kaga-city-iot-33.html

ICTに関する人材育成政策と人材

◆ICTに関する人材育成政策と個々の組織・個人の自助努力が重なれば、ICTの利活用は一層進展します。





- 個々の組織・個人が自助努力をすることで、
  人材育成政策と自助努力のサイクルも、ICT・データの利活用のサイクルと同様に回していきましょう。