

デジタルスキル標準
DXリテラシー標準
DX推進スキル標準

社会におけるデータ

Why DXの背景

社会の変化
顧客価値の変化
競争環境の変化

What DXで活用されるデータ・技術

データ	社会におけるデータ
	データを読む・説明する
	データを扱う
	データによって判断する
デジタル技術	AI
	クラウド
	ハードウェア・ソフトウェア
	ネットワーク

How データ・技術の利活用

活用事例・利用方法	データ・デジタル技術の活用事例
	ツール利用
留意点	セキュリティ
	モラル
	コンプライアンス

マインド・スタンス

デザイン思考／アジャイルな働き方 新たな価値を生み出す 基礎としてのマインド・スタンス	顧客・ユーザーへの共感	常識にとらわれない発想	反復的なアプローチ
	変化への適応	コラボレーション	柔軟な意思決定

この教材の学習目標と学習項目

「データ」には数値だけでなく、文字・画像・音声等様々な種類があることや、それらがどのように蓄積され、社会で活用されているか知る。

- ビッグデータの活用が進んでいる中で、機械の稼働状況のようなデータだけでなく、文字、音声、動画や、自分の行動履歴もデータとなりうることを理解することが求められる。
- ビッグデータを得体のしれないものとして恐れるのではなく、便利に利用するために、ビッグデータがどのように使われているのか知る必要がある。

DXリテラシー標準学習項目例

- データの種類
 - 取得方法による分類
 - 取得主体による分類
 - データそのものの属性による分類
- 社会におけるデータ活用
 - ビッグデータとアノテーション
 - オープンデータ

GIF追加学習項目

- データの重要性
- データと情報と知識
- データの信頼性と品質
- データの蓄積
- セキュリティ

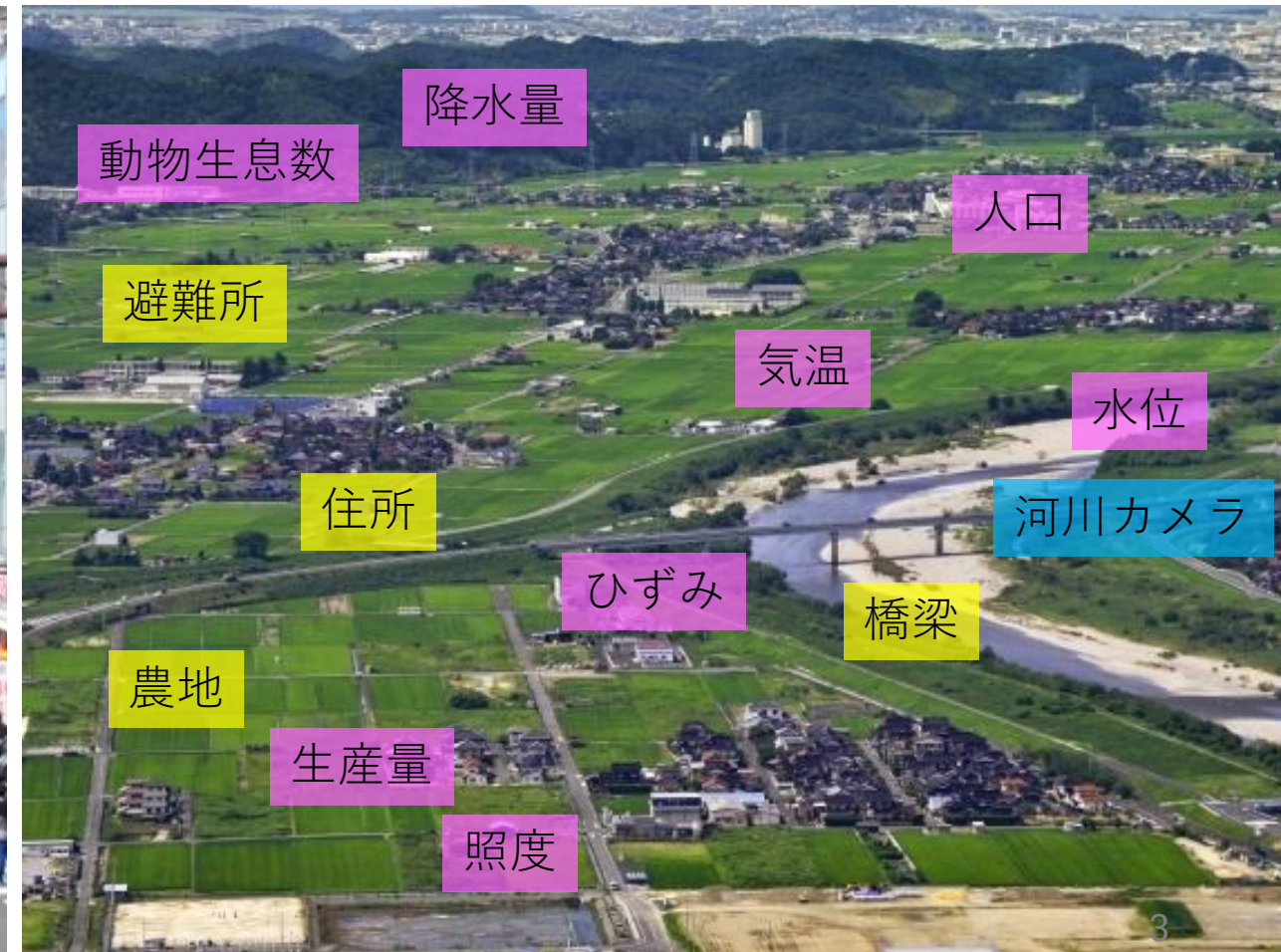
社会はデータであふれています

これらをつなぐと新たな価値が生まれてきます。

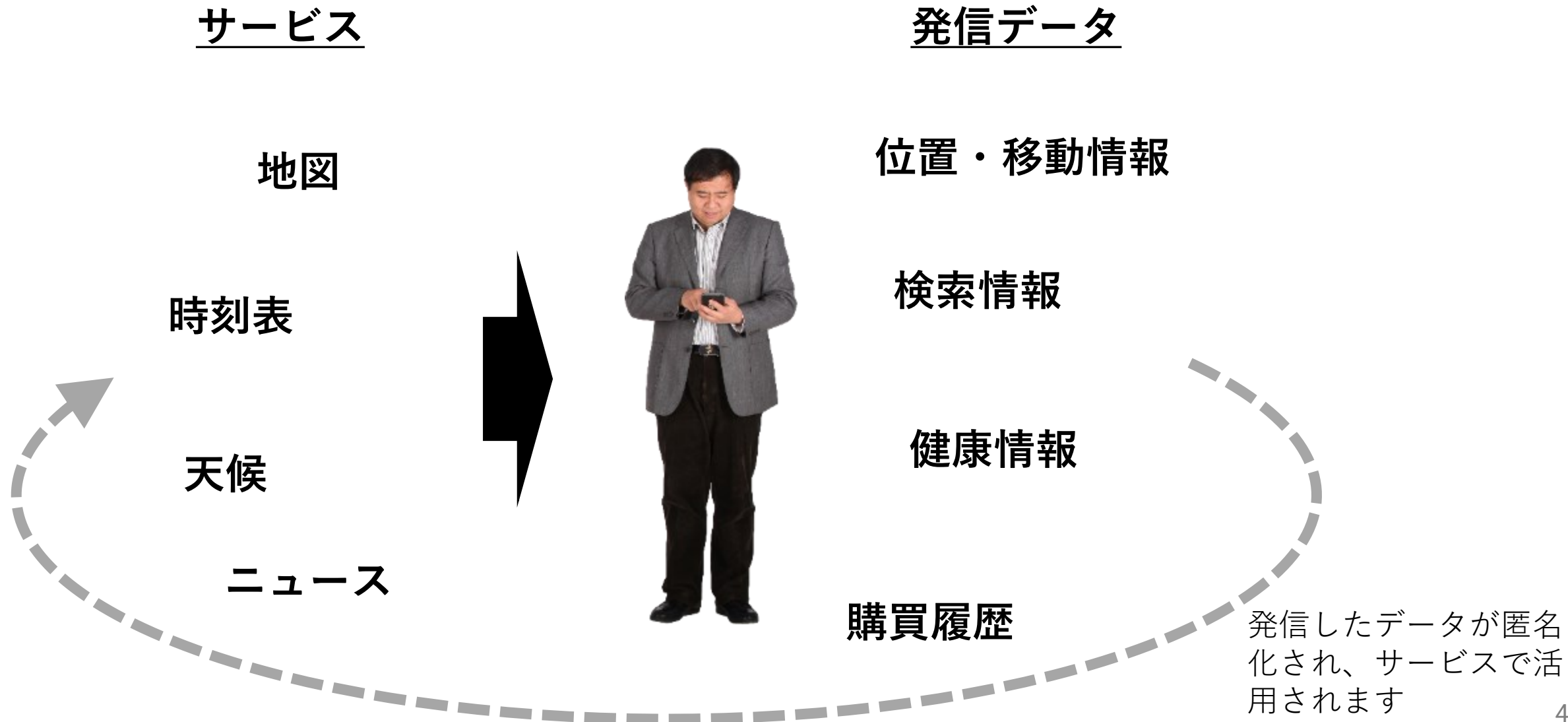
文字データ

数値データ

画像データ



誰もがデータ消費者であり、データ生成・発信者です

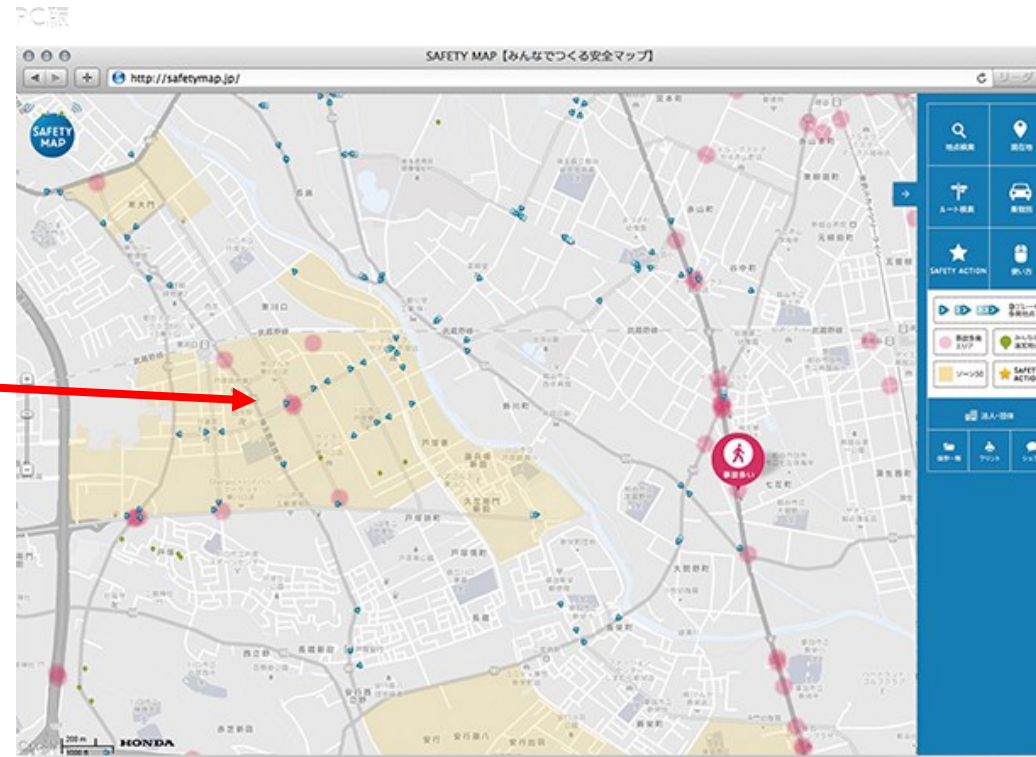


今まで気にしていなかったデータが価値を生みます

- 車の運転データを集積することで、危険箇所がわかるようになります。



急ブレーキ
急ハンドル情報



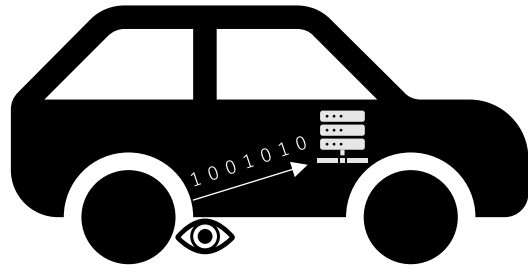
スマートフォン版



データは様々なものにつながります

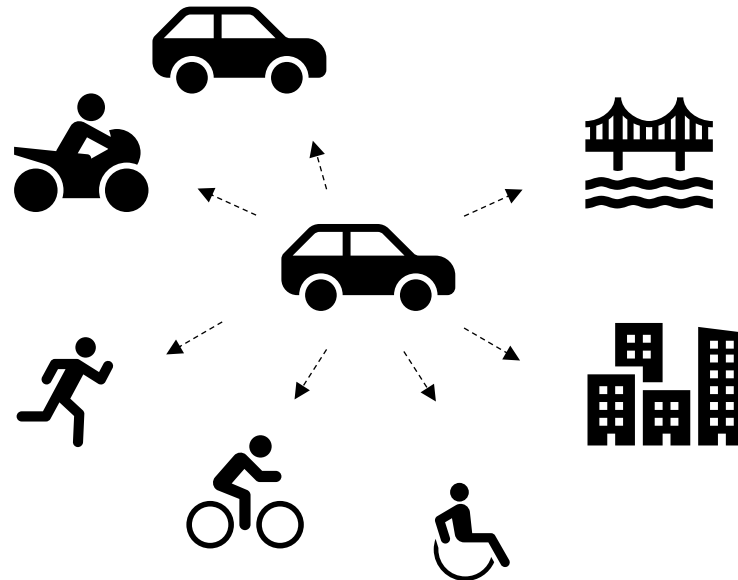
- 「もの」を最高の状態にするためにセンサーデータを使うだけでなく、そのデータが「様々なもの」につながり新たな価値を生み出します。

以前



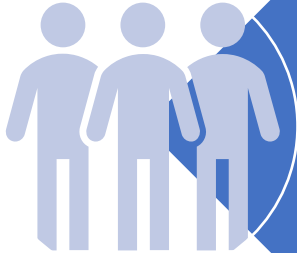
路面の状態のデータが
運転支援に使われる
＝車でデータが完結

現在



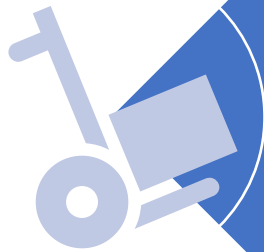
路面のデータを、他の乗り物や都市管理で使うことで、各サービスが使いやすく、社会全体が暮らしやすくなっています

データにより人、物、金の流れが変わります



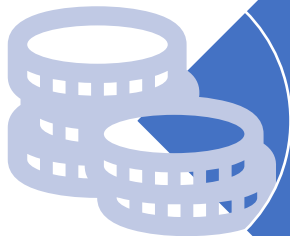
人流

- 混雑予想を公開することでピークを変える
- クーポン発信で人流が変わる
- テレワーク、Web会議で働き場所や移動が変わる



物流

- ネットショップで個配が増える
- 自動配送は配送の体制を大きく変える



金流

- キャッシュレスで事務作業が減る
- クラウドファンディングで資金集めができる

データの基礎

何故データが重要なのでしょうか

- 経営でも機器の活用でも、データを使わずに何かを実現することはできません。
- また、データは寿命が長く、波及範囲が広いです。

業務プロセス
寿命1年

プロセスをノウハウ化して共有する仕組みが必要

新技術やユーザーを見て随時変更

変更が可能なシステムが必要

アジャイルとDevOpsが前提

情報システム
寿命10年

システム標準化はデータとインタフェースの標準化が重要

リプレイス前提

ビルディングブロック化

汎用サービスの活用

データ
寿命100年

組織のアンカーとしての機能

外部との組み合わせで価値が増大

様々な活用が可能なりソース

利活用ルールの見直しが必要

データって何でしょう

- データには様々な観点があります。

種類

- どのようなデータなのか

構造等

- 使いやすいデータなのか

取得方法

- どうやって入手したのか

取得主体

- 誰から入手したのか

個人情報

- 配慮が必要な情報なのか

データの種類（分類）

- 文字
 - －文字による文書やSNS、表等のデータ。
- 音
 - －会議録、騒音等のデータ。
- 画像
 - －商品画像、SNS、監視カメラ等のデータ。
- 動画
 - －広報、セミナー等のデータ。字幕データがついているものもある。
- 数値データ（センサーデータ）
 - －センサーの計測値などのデータ。

データの種類（構造等）

- 構造化データ

- 主に文字や数値のデータです。
- データが「タイトル」「日付」「概要」「場所」「連絡先」のようにデータ項目の組み合わせでできています。
- 入力確認や補助が行いやすく、データの活用、AI等による解析も行いやすいデータです。

- 非構造化データ

- 文章、音声、画像、動画等のデータで、データを解析しないと高度な活用が難しいデータです。
- データ全体の80%が非構造化データと言われています。

- メタデータ

- 「データ名称」「作成日」等、データを解説するための見出し情報のデータです。データの検索用に使います。

データの種類（取得方法）

- 行動ログデータ
 - － 人が行動することで、購買や健康等の様々なデータが蓄積されます。誰がどのデータにアクセスしたかというアクセスログも行動ログの一種です。
- 機械の稼働ログデータ、センサーデータ
 - － 機器や都市に設置されたセンサーから自動収集されるデータです。一定間隔で取得するデータ、障害発生時に取得するデータ等があります。
- 実験データ、調査データ
 - － 実験や調査の目的に応じて収集されるデータです。ウェブフォームを使ったアンケート等があります。調査対象の年齢や地域の分布、調査数等、調査結果が偏らないような配慮が必要です。
- 生体データ
 - － 指紋や虹彩、顔画像、声、健康情報といった人体に由来するデータです。本人確認や健康管理などに用いられます。

データの種類（取得主体）

- 一次データ

- ー 情報源から直接取得したデータです。エラーデータ等を含む場合があります。

- 二次データ

- ー 他社によって情報の編集やクレンジングされた使いやすいデータです。

- ・ 一次データ等の原データの編集方法、取得方法、中間関与者の情報が求められます。（プロビナンス情報）

➤ 一次データは、整理されておらず処理が大変ですが、エラーの原因や少数意見等、中間者が省略してしまう情報を含んでいる場合があります。二次データを使いながら、一次データを必要に応じて参照するなど、目的に応じてデータを使い分けます。

データの種類（個人情報）

- 個人を特定できる「氏名」等の情報です。
- 複数のデータを組み合わせることで、個人が特定できてしまうため、個人情報になり得る情報というものもあります。
- 個人の情報を使うことで、利用者の嗜好に合わせた情報を提供したり、手続きでの入力を自動化したりすることができます。利便性と安全性を考えて、利用者の同意のもとで活用することが求められています。
 - － 最近では、同意のもとで氏名や生年月日、住所、決済情報などの個人情報をも民間サービスに預ける人が増えています。健康情報を提供するケースも少なくありません。

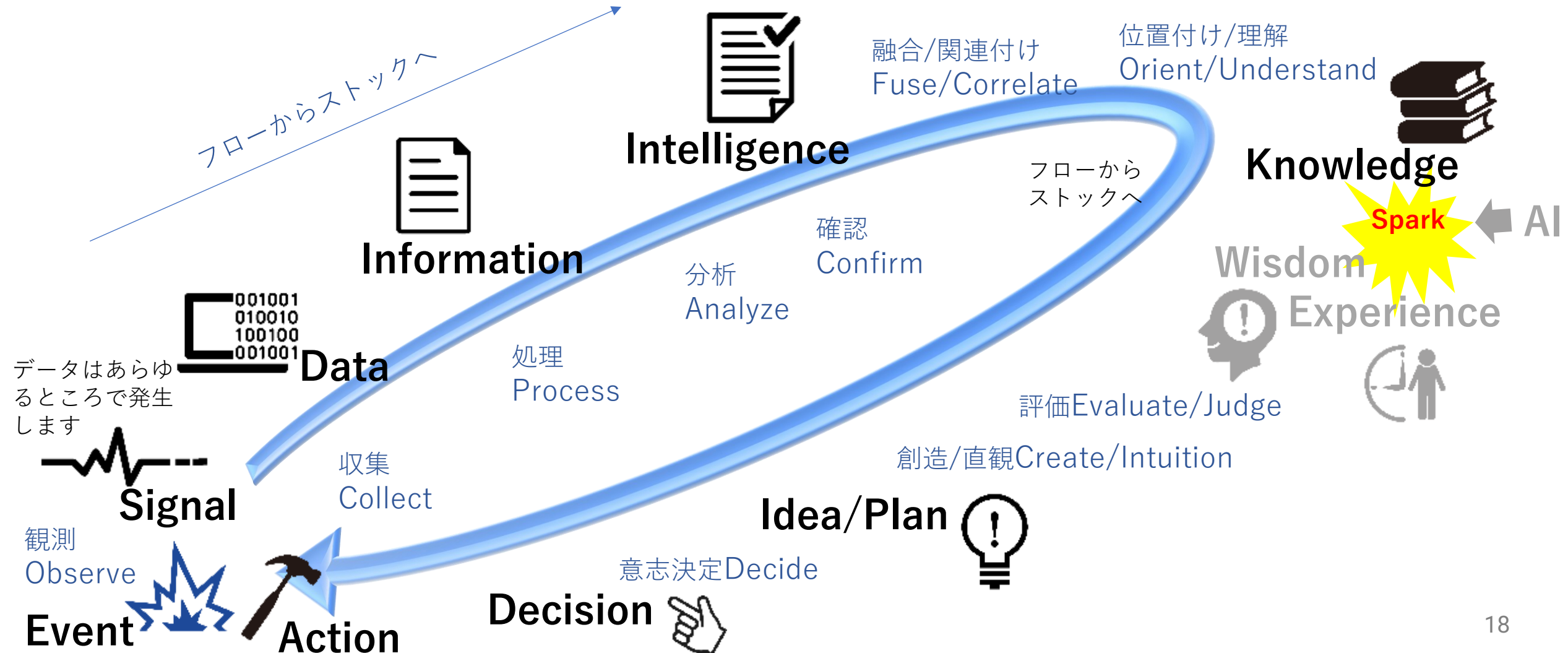
地理空間データとセンサーデータ

- データは、取得地点の情報などの地理空間データに紐付けられて使うことが多いです。
 - 地理空間情報は、地図データ、緯度経度などのデータ、敷地などの形状、空間情報、建物情報などを含みます。
- センサーデータは、センサーなどのデバイスから取得するデータです。そのデータをデバイス上で処理してアクチュエータを動かす場合や、データをエッジ上のAIで処理して匿名化したうえで送信することもあります。
 - エッジデバイスは屋外等の条件が厳しい環境に設置されることも多く、データの補正処理などが必要になるケースもあります。

社会におけるデータの活用と管理

データのサイクルを見てみましょう

データを知識にして、更に、新たなものを生み出せるサイクルが必要です。

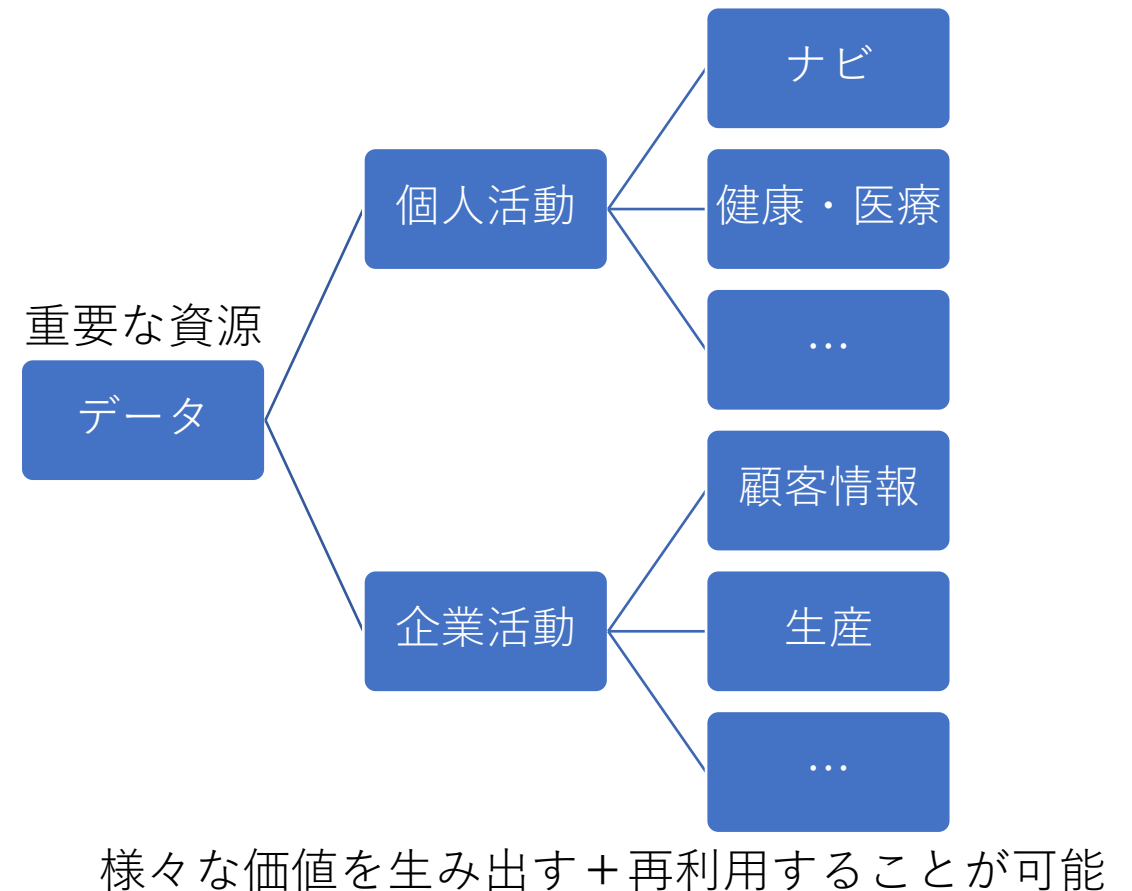
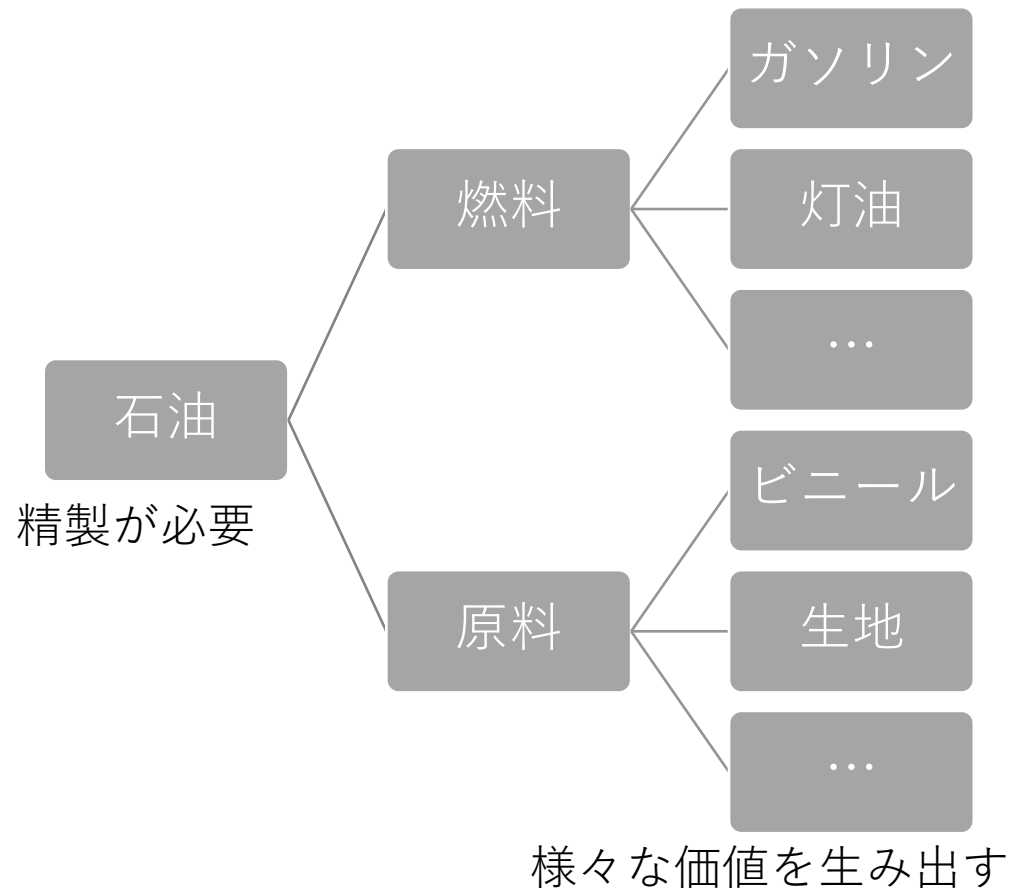


データの量が急速に増えています

- 調査会社IDCが2018年に公表した「Data age 2025」では、2018年に33ZBのデータ量が2025年には175ZBになると予想しています。
 - 2025年には、90ZBはIoTデバイスでデータが送出されます
 - 2025年には、ほぼ30%のデータがリアルタイムデータになります
- 画像データ、センサーデータ、3Dデータ等、これまで手に入らなかったデータの洪水がやってきます。
- これらのデータの活用をできるかどうかが組織の競争力の鍵になります。

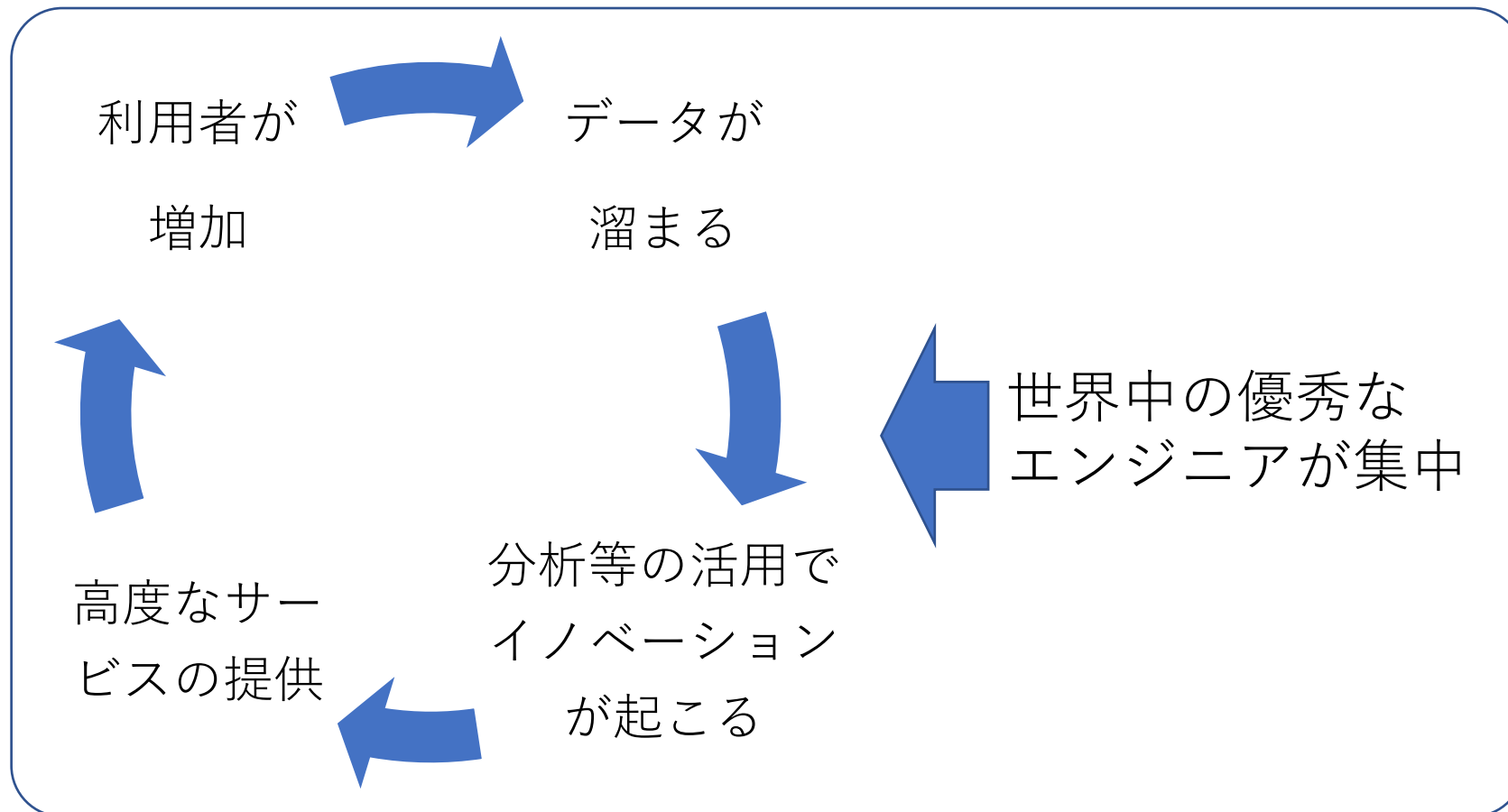
データは21世紀の石油と言われています

- 精製・加工することで、様々な価値を生み出してくれます。



データ収集によるスパイラルアップを目指します

- データは量が増すことで力が増していきます。また、複数のデータを組み合わせることで、新しい価値を生み出すことができます。



この動きがグローバルに展開している

データを活かす技術

- データは分析や可視化して使われます。特にAIやビッグデータ解析があります。
 - ー ビッグデータ解析
 - ・ 大量のデータを解析することで、データの相関性等から、少量のデータでは見つけられなかった事実を発見します。
 - ・ 販売データ、会計データ、機械振動、SNS等のさまざまなデータが解析に使われています。
 - ー AI
 - ・ データからパターンを抽出・学習し、新たな入力に対して予測することが出来るツールです。
 - ・ 画像やテキスト等の非構造化データの取り扱いも得意とします。

※ビッグデータ解析やAIを使いこなすためには、大量のデータや質の高いデータが必要になります。

データの設計

- データは、様々なサービスで使われることも多いため、相互運用性（インタオペラビリティ）を確保した設計にしておく必要があります。また、匿名性やセキュリティを確保しやすい設計が採用されることもあります。
- データの設計にリファレンスモデル（ひな形）を使うことで、データ設計が迅速にでき、しかもインタオペラブルな設計ができるようになります。
 - ー 日本では、政府相互運用性フレームワーク（GIF）のデータモデルが提供されています。
- また、データの意味を正確に定義するために、データ辞書を使ったり、既存のコードリストを活用することも有効です。

データの収集方法

自分の保有するデータを活用するだけでなく外部データを組み合わせるデータを活用することがあります。

行政機関の統計データに加え、統計以外のオープンデータや民間企業が販売するデータの活用が増えています。

- 組織内データの活用

- － 組織内でどのようなデータを保有しているのか、棚卸しを行います。

- ・ このとき、データの管理に標準的なメタデータを使うことで、管理がしやすくなり、再利用性が高まります。

- オープンデータの活用

- － 行政機関が保有する各種届け出や、一覧化したデータ等を誰でも使えるように公開しています。

- － 今後は、法人や土地情報等、デジタル社会の台帳と言われるベース・レジストリの公開も進んでいくものと期待されています。

リアルタイム・モニタリング

- 工場でセンサーデータをリアルタイム・モニタリングすることは当たり前のように行われていますが、都市の人流データや窓口の処理データなどもリアルタイムで収集、可視化できるようになってきています。
- 社会調査は、年1回で行われるものも多いですが、デジタル技術の発達により、データの直接計測や、レジ情報の収集、フォーム入力による回答など、リアルタイムで情報を収集しやすくなってきています。
- 既存の方法にとらわれない、モニタリング方法を検討していく必要があります。

データにおける信頼と品質

- 外部データを使用する場合には、そのデータの信頼性が重要になります。
 - ー 以下のような観点での確認が必要になります
 - ・ そのデータにエラーデータがどのくらい含まれるのか
 - ・ データの精度は十分か
 - ・ 必要な粒度が確保されているか（粒度とは、分、時間、日等のデータの取りまとめ範囲）
 - ・ 原データを提供している関係者やその手法は信頼できるのか
 - ・ データの転送や加工は適正な方法で行われているか
- 特に精度や品質は、データの取引価格に関係します。また、十分な品質が確保されないデータを手に入るとデータ活用する前にデータのクレンジング作業が必要になるので、取引を検討するときに確認が必要です。
- データサイエンティストが不足していると言われます。一方、データサイエンティストの7-8割の仕事はデータの分析ではなく事前のクレンジング作業をしています。品質の高いデータを手入手してこの工程を省けるようにすることが重要です。

データの蓄積

- データは、個人で管理したいもの、安全に管理したいもの、複数人で共有して管理したいもの等があります。目的に応じて蓄積方法を選択します。
 - －クラウド（占有領域）
 - ・ 個人の写真や文書等を保管するためのネットワーク上の保管領域です。利用者の認証により、PCや携帯端末など複数の端末からアクセスできるサービスがあります。
 - －クラウド（共有）
 - ・ 組織内でデータを共有したり、複数人でデータを共有するためのネットワーク上の保管領域です。文書を複数人で同時編集するなどの作業が可能なものもあります。
 - －組織内サーバー
 - ・ 組織内や外部委託者の施設に設置したサーバーに保管します。電源やデータのバックアップ装置等が必要になります。
 - －自己端末
 - ・ PC等の自分の端末内にデータを蓄積します。
 - －エッジ
 - ・ センサーは、データを蓄積しないものが多いですが、自端末でデータを処理したり、データを一時的に記録をしてサーバーに一括転送するもの等があります。

データのセキュリティ

- データの漏洩や侵入による盗難は、直接的な被害以外にも組織イメージの低下などのダメージが大きいです。
 - ー 最近では、侵入を防ぐ壁（ファイアウォール）型の防御だけでなく、内部の不審操作のモニター等、多様なセキュリティ技術が提供されています。
 - ー 一方、人的要因によるデータの流失なども多く、社員教育等が重要になります
- 最近、サーバーにロックをかけてデータやシステムを操作できないようにして身代金を要求するランサムウェアと言われる犯罪が増えています。
 - ー バックアップを複数拠点に保管するなど、業務停止にならない対策をする必要があります。

まとめ

社会には使われていない多くのデータがある

組織にとって、データを活用して価値を生み出すことが重要である

データは集めれば良いというものではなく、種類や構造を理解して使うことで大きな効果を引き出せる

データはきちんと設計、蓄積・管理される必要がある

改訂情報

- 2023-03-31
 - － デジタル庁がGIFアカデミーとして公開
- 2025-07-23
 - － GIFアカデミーの資料を元に以下の内容を追記し、他教材とともにシリーズ化
 - ・ 地理空間情報とセンサーデータの追加
 - ・ データの設計の追加
 - ・ リアルタイム・モニタリングの追加