

行政機関におけるデータ活用

九州テレコム振興センター
井上 英幸

【次第】

- 1. データ利活用の基本概念**
- 2. 世の中の動き(国の政策や民間の動向)**
- 3. オープンデータ推進の現状と課題**
- 4. データ活用のポイント**

1. データ利活用の基本概念

1.1 データ活用の時代

注目されるIT技術

IoT, AI, Bigadata

行政分野の新しい流れ

EBPM

国の動き

官民データ活用推進基本法

(世界最先端デジタル国家創造宣言

・官民データ活用推進計画)



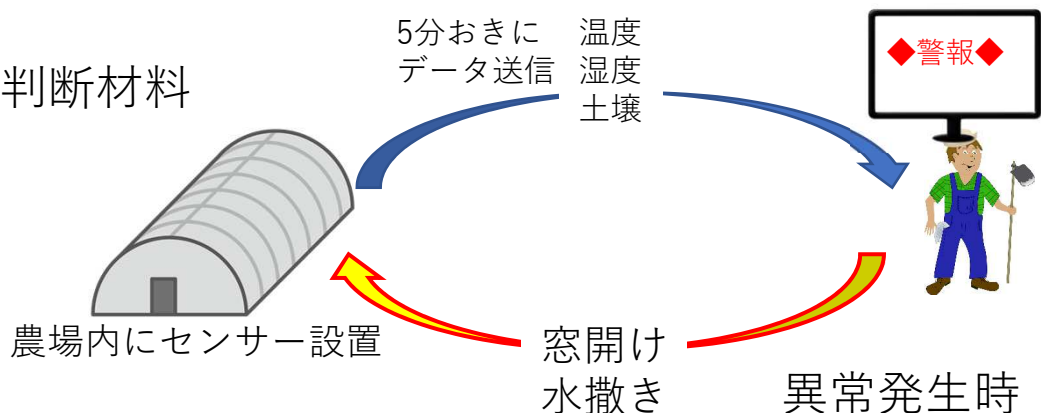
共通するのは

データを
活用すること

1.2 データ活用の例(農業分野)

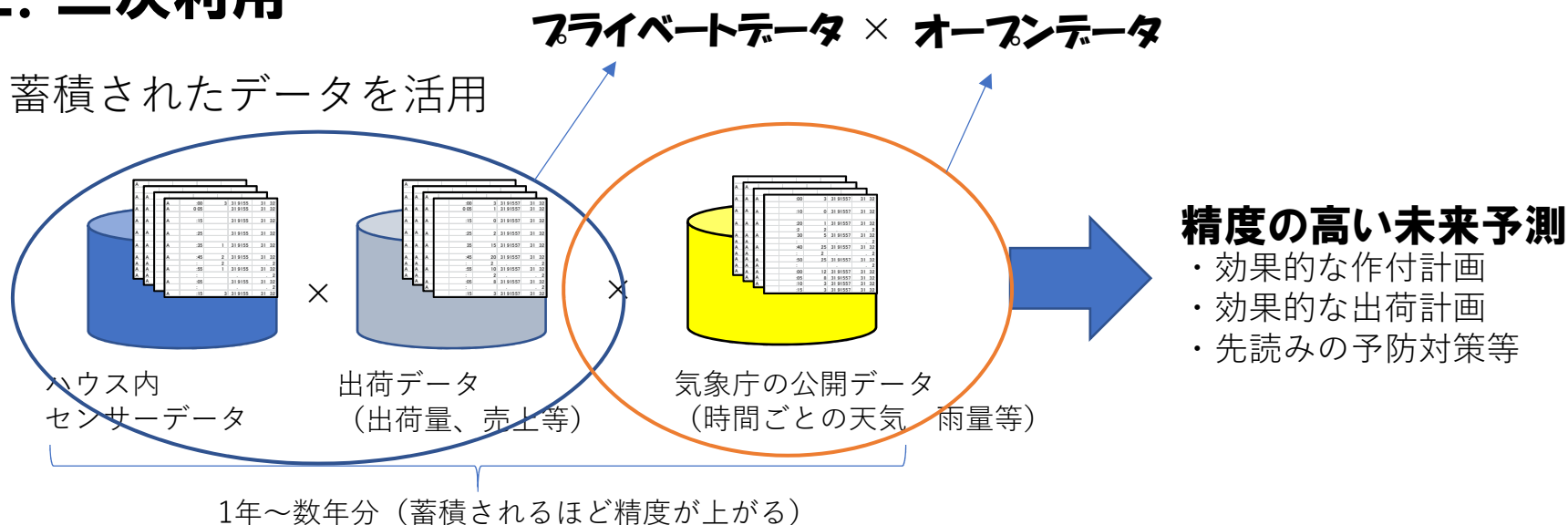
1. 一次利用

リアルタイムな判断材料



2. 二次利用

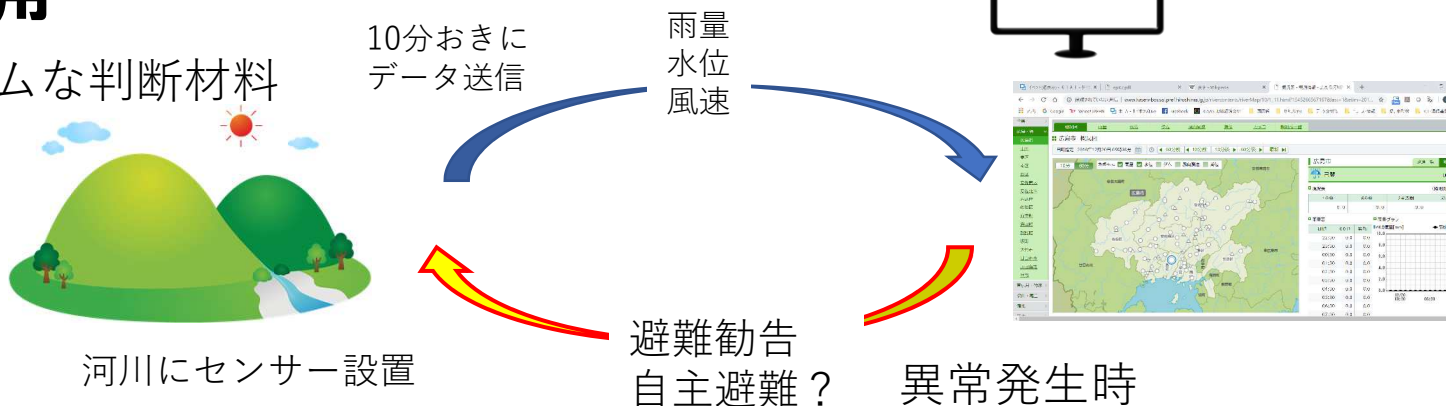
蓄積されたデータを活用



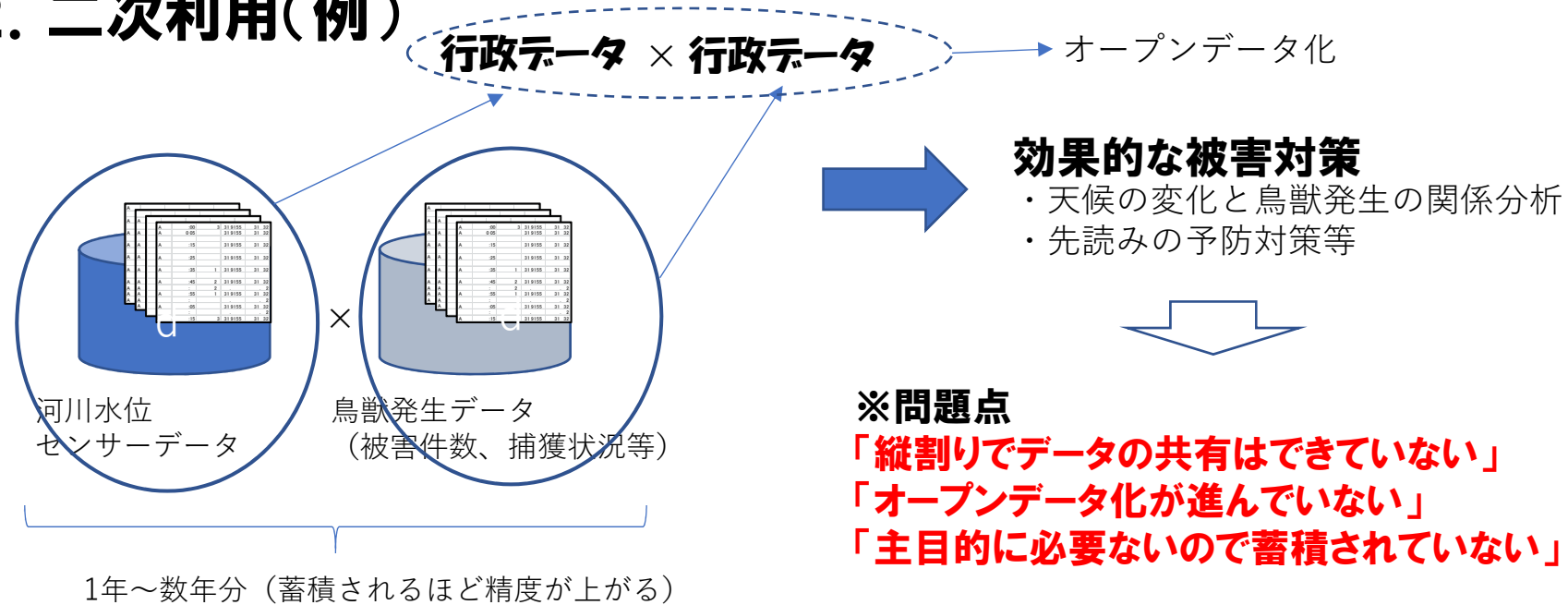
1.3 データ活用の例(仮:行政(土木)→行政(農政))

1. 一次利用

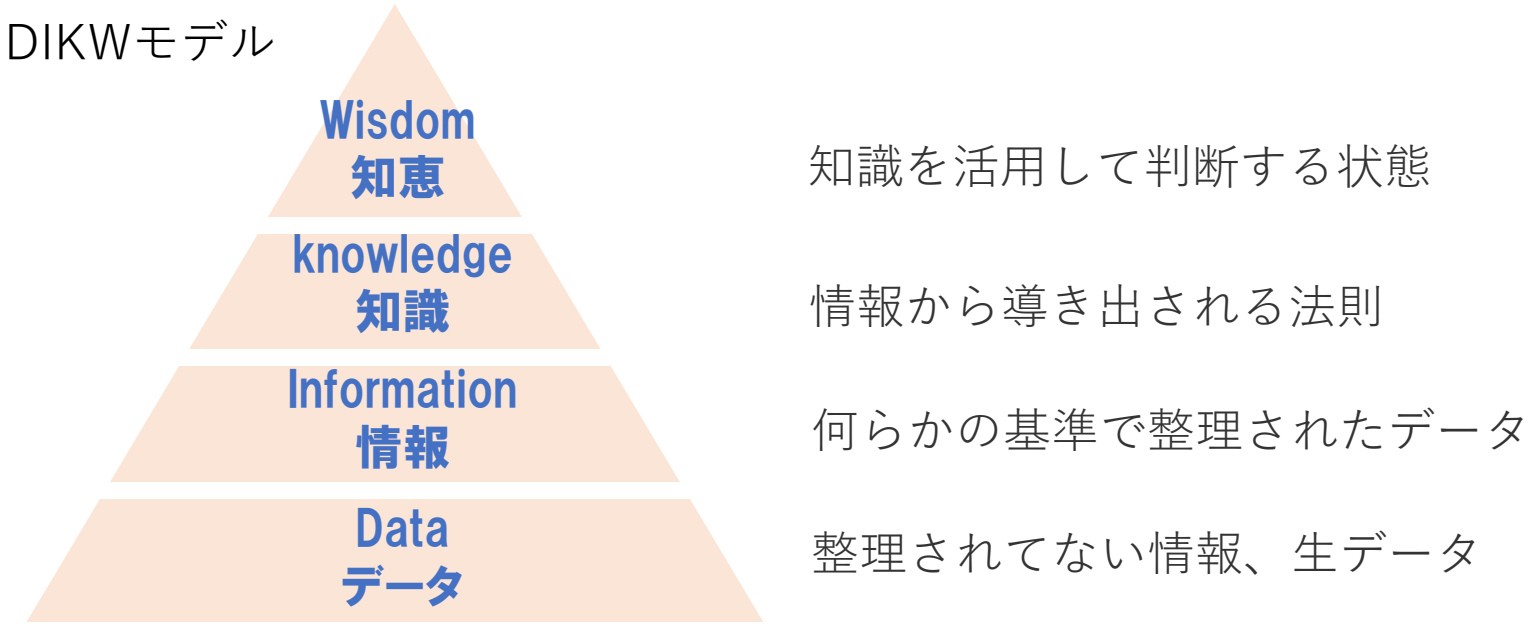
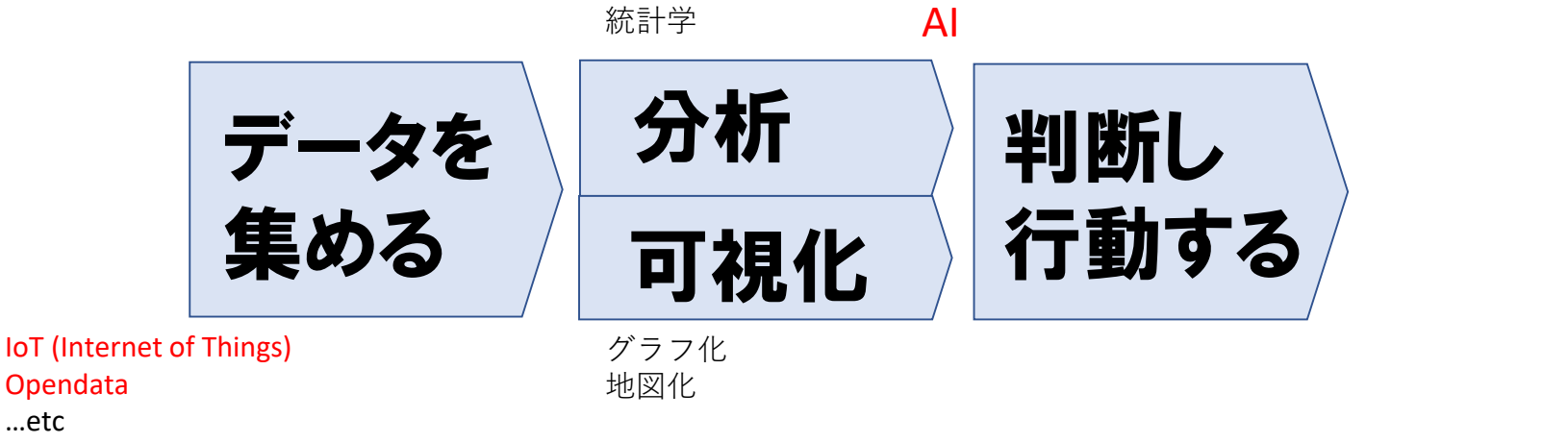
リアルタイムな判断材料



2. 二次利用(例)



1.4 データ活用の流れとDIKW



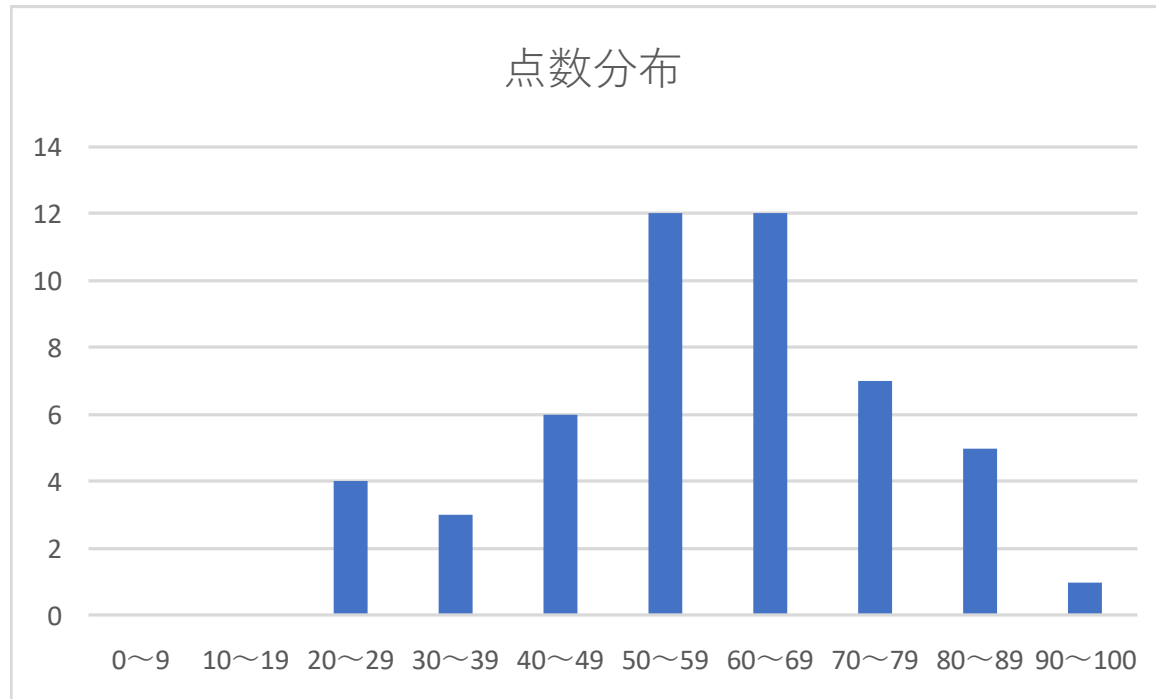
あるクラス（50人）の試験の点数

53	36	46	52	21
69	19	17	56	45
36	76	56	72	72
28	31	71	45	67
64	47	45	76	46
66	49	70	46	66
53	60	18	40	42
40	53	54	28	53
42	87	41	58	52
35	48	56	65	20

あるクラス（50人）の試験の点数 ~合格点70以上を黄色表示

53	36	46	52	21
69	19	17	56	45
36	76	56	72	72
28	31	71	45	67
64	47	45	76	46
66	49	70	46	66
53	60	18	40	42
40	53	54	28	53
42	87	41	58	52
35	48	56	65	20

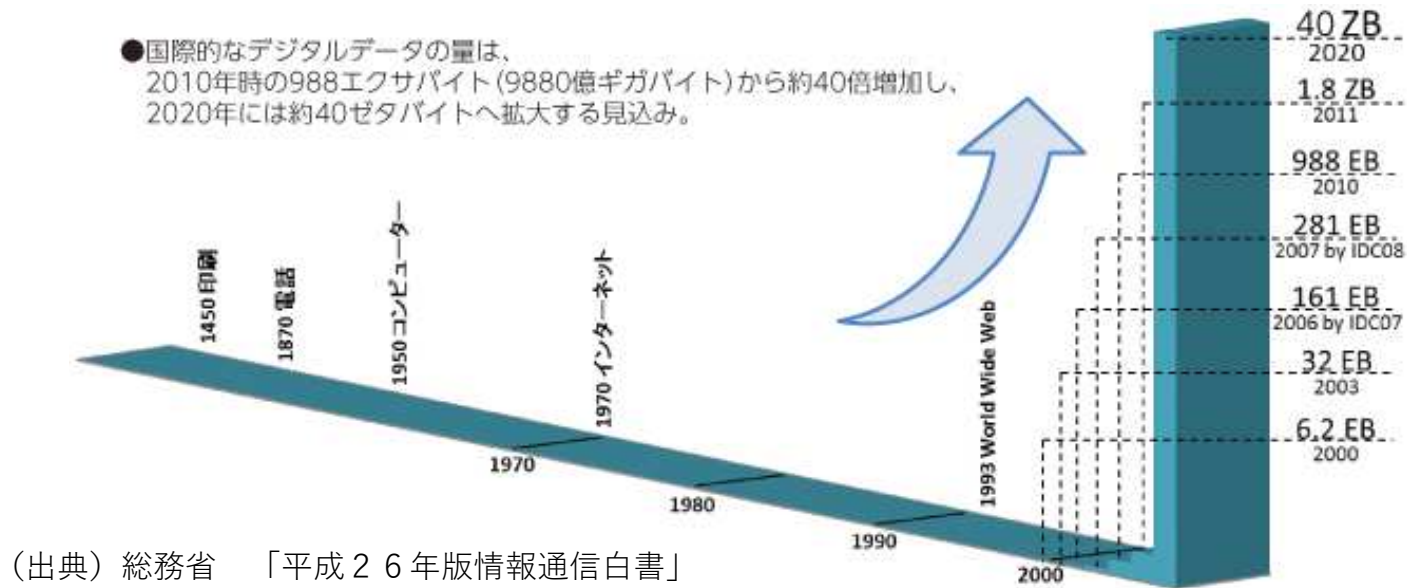
ヒストグラム化、基本統計量の算出（平均、標準偏差、中央値、最頻値・・・）



基本統計量	
平均	49.76
標準誤差	2.368942
中央値（メジアン）	50.5
最頻値（モード）	53
標準偏差	16.75095
分散	280.5943
尖度	-0.41807
歪度	-0.1292
範囲	70
最小	17
最大	87
合計	2488
データの個数	50

2. 世の中の動き(国の政策や民間の動向)

2.1 爆発的に増えるデータ



**パーソナルデータは
インターネットにおける新しい石油であり
デジタル社会における新たな通貨である**

世界経済フォーラムが2011年に公表した報告
「パーソナルデータ：新たな資産カテゴリーの出現」より

2.2 トピックス(データの寡占化への懸念)

朝日新聞デジタル > 記事

巨大ITの規制、政府強化へ データ集中・市場独占懸念

西山明宏 2018年11月6日01時49分

シェア 81 ツイート list 13 ブックマーク メール 印刷

政府の巨大IT規制のイメージ

規制対象	G Google A Apple F facebook A amazon	など巨大IT企業
問題点	取引などのルールが不透明	特許や個人情報などデータ独占
検討される主な規制	専門家による監視組織 ルールの開示義務づけ	独禁法(優越的地位の乱用)の適用 企業結合の審査時にデータ集積の度合いを判断材料に

政府は米 Google や 米アマゾン など、「プラットフォーム」と呼ばれる巨大IT企業への規制を強化する。個人情報や知的財産などのデータがこうした企業に集中して市場の寡占が進み、公正な競争環境をゆがめかねないためだ。欧州連合(EU)などの規制強化の動きと歩調を合わせる。

7月から議論を進めてきた 経済産業省 や 公正取引委員会 などの検討会が5日、中間論点整理案を公表した。年内に最終案をとりまとめ、年明けから具体的な規制方法の検討に入る。



(参考)
EU ~ **GDBR** (一般データ保護規則)

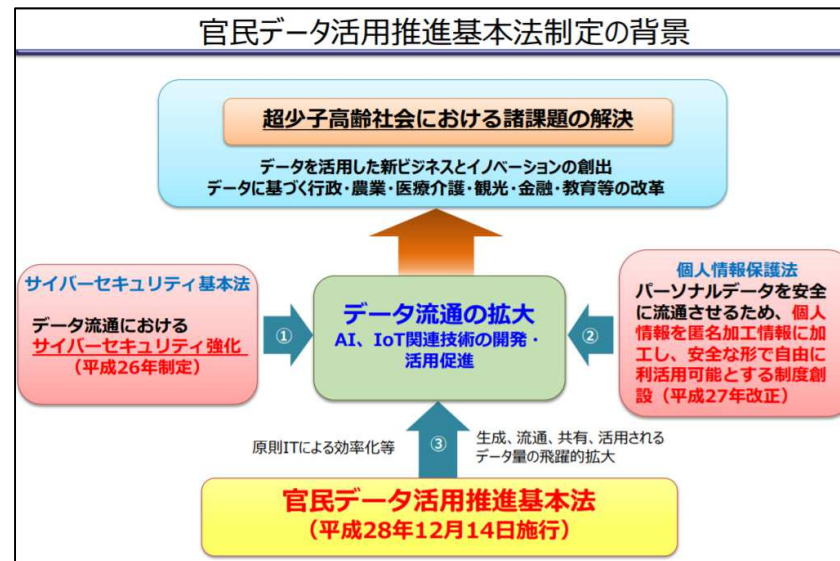
2016年5月24日に発効、2018年5月25日から適用開始
個人データのEEA(欧州経済領域)“域外”への持ち出しは原則禁止。違反者には高額な罰金が科せられる

2018年11月6日 朝日新聞DIGITALサイトより

2.2 データ主導社会へ向けて

2016.12 官民データ活用推進基本法公布・施行

2017.05 改正個人情報保護法全面施行

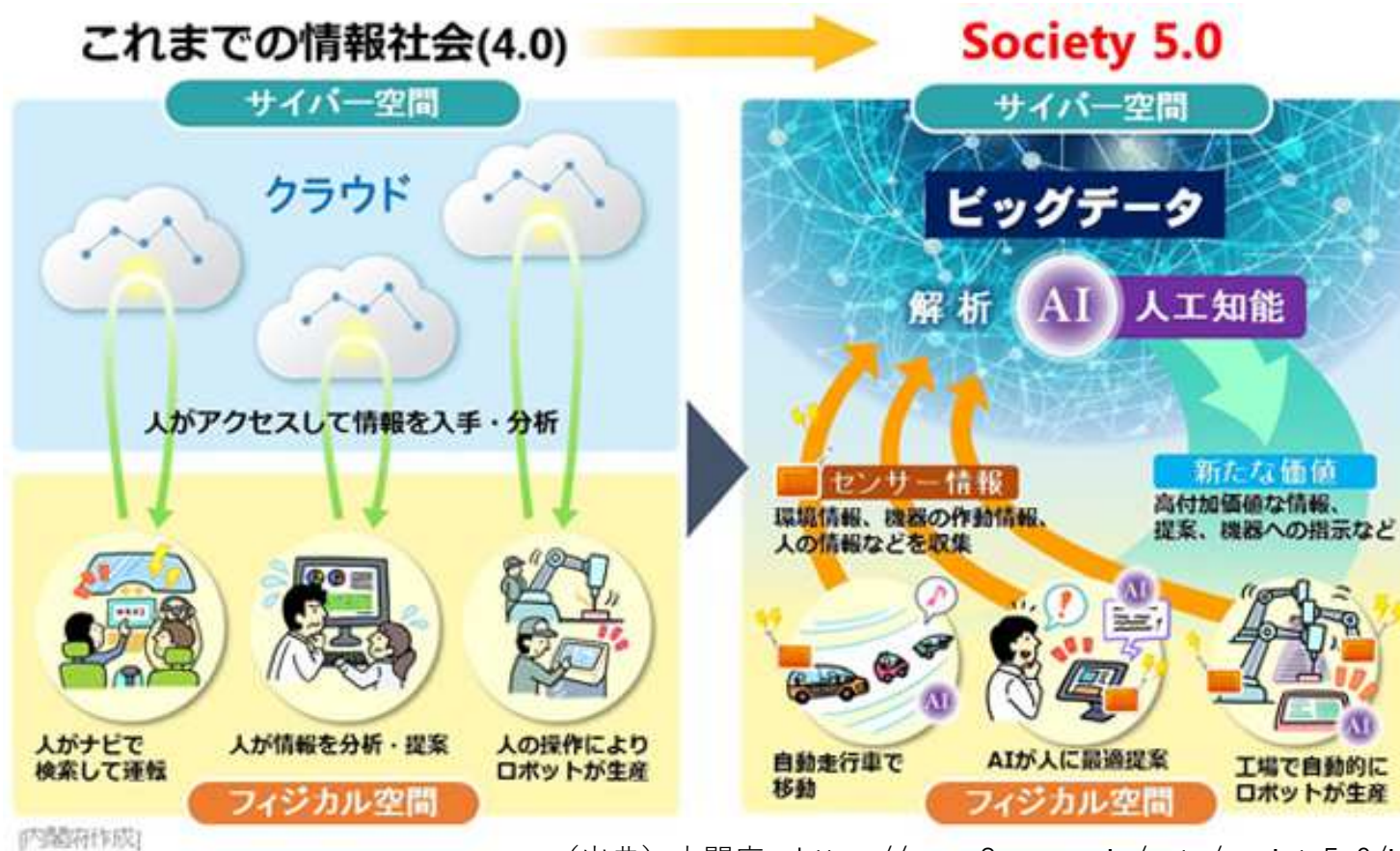


2017.05 オープンデータ基本指針

公共データにつきオープンデータを前提として情報システムや業務プロセス全体の企画、整備及び運用を行う「オープンデータ・バイ・デザイン」の考え方にに基づき、政策企画・立案の根拠となったデータを含め、各府省庁が保有するデータはすべてオープンデータとして公開すること、各省庁のウェブサイト公開されるデータについては政府標準利用規約を適用し公開データの二次利用を積極的に促進すること、機械判読に適した構造・データ形式で掲載することなどの原則を提示

2.3 Society 5.0

内閣府の第5期科学技術基本計画において、我が国が目指すべき未来社会の姿として提唱されたもの



(出典) 内閣府 https://www8.cao.go.jp/cstp/society5_0/index.html

サイバー空間（仮想空間）とフィジカル空間（現実空間）を高度に融合させたシステムにより実現

IoTとAIにより劇的な社会変化が生まれる！

3. オープンデータ推進の現状と課題

3.1 オープンデータの定義

国、地方公共団体及び事業者が保有する官民データのうち、国民誰もがインターネット等を通じて容易に利用(加工、編集、再配布等)できるよう、次のいずれの項目にも該当する形で公開されたデータをオープンデータと定義する。

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

〈参照:オープンデータ基本指針(平成29年5月30日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決定)〉

3.2 オープンデータの意義

1. 国民参加・官民協働の推進を通じた諸課題の解決、経済活性化

広範な主体による公共データの活用が進展することで、創意工夫を活かした多様なサービスの迅速かつ効率的な提供、官民の協働による公共サービスの提供や改善が実現し、ニーズや価値観の多様化、技術革新等の環境変化への適切な対応とともに、厳しい財政状況、急速な少子高齢化の進展等の我が国が直面する諸課題の解決に貢献することができる。

また、ベンチャー企業等による多様な新サービスやビジネスの創出、企業活動の効率化等が促され、我が国全体の経済活性化にもつながる。

2. 行政の高度化・効率化

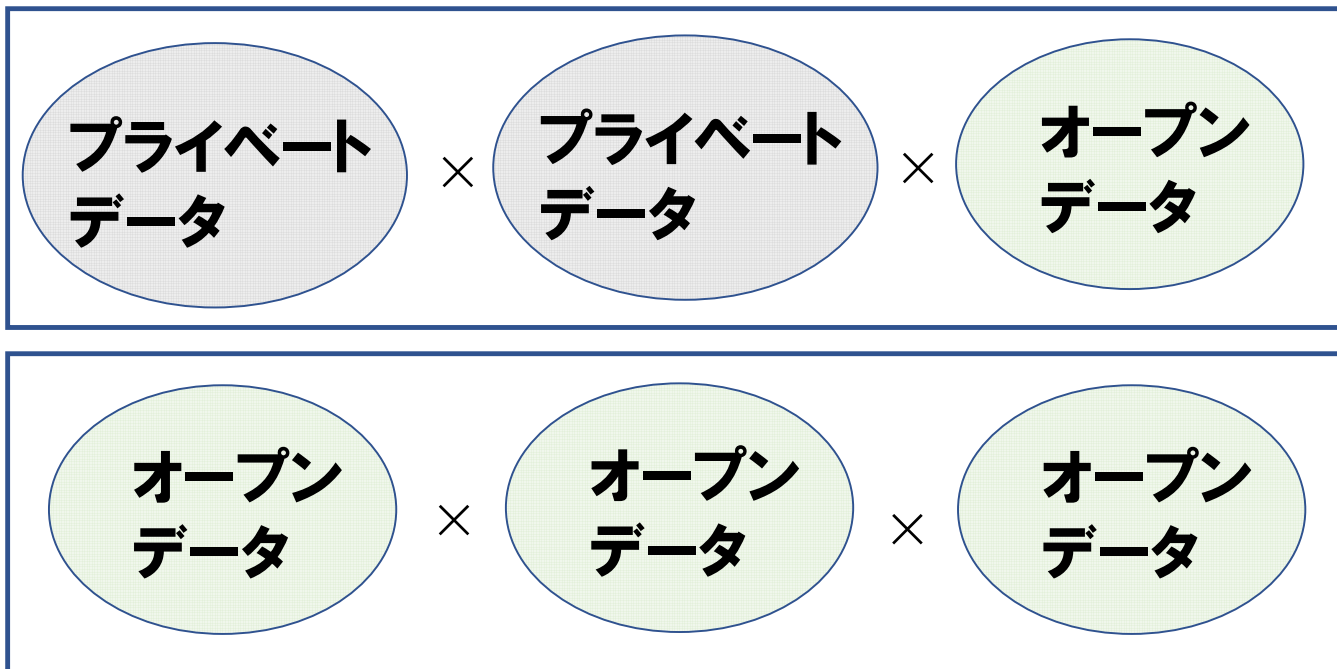
国や地方公共団体においてデータ活用により得られた情報を根拠として政策や施策の企画及び立案が行われることで（EBPM：Evidence Based Policy Making）、効果的かつ効率的な行政の推進につながる。

3. 透明性・信頼の向上

政策立案等に用いられた公共データが公開されることで、国民は政策等に関して十分な分析、判断を行うことが可能になり、行政の透明性、行政に対する国民の信頼が高まる。

3.3 オープンデータの利用価値は？

オープンデータで何がどう変わるのか？と問われた際に、多くの自治体担当者が悩むケースが多いが、1つのオープンデータだけを使う方が稀で、様々なデータを組合せて分析するケースが一般的だと考えられる



3.4 オープンデータの取組状況(地方自治体)

※ここでいう「取組」とはオープンデータとして最低1つでも公開しているかどうかを指す

平成30年9月17日時点

オープンデータ取組状況 (都道府県)

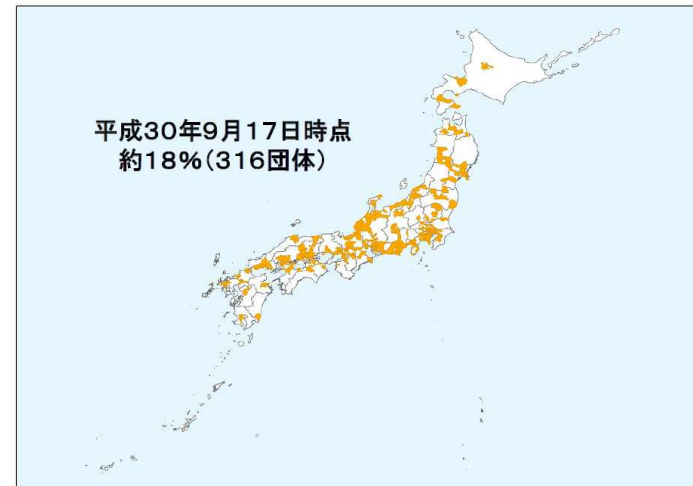
※平成30年4月30日時点



出典: 政府CIOポータル「オープンデータ取組済自治体一覧(平成30年4月30日時点)」をもとに作成

オープンデータ取組状況 (市区町村)

※平成30年9月17日時点



出典: 政府CIOポータル「オープンデータ取組済自治体一覧(平成30年9月17日時点)」をもとに作成

都道府県は (一応) 100% (47団体)

市区町村は 18% (316団体)

ただし、現状は総じて中身が薄い状況
(インターネット草創期の自治体ホームページの開設状況に近い?)

3.5 オープンデータ～実際の取組で見えてきた課題

宮崎県庁在籍時、平成26年度に行った保有データの棚卸し

調査の概要

調査期間：平成26年10月～12月（出先は平成27年1月～2月）

調査方法：各所属毎に調査票を配布

データごとに1行で概要を記載させる（15項目程度）

※データそのものは集めておらず、概要の調査

平成26年度の庁内所有データの調査結果

全調査対象：本課 89（回答89課…100%），出先機関 107（回答60課…56%）

総データ数：約3,600件（1所属平均 約24件）

公開可否状況：公開可能または一部公開可能 53%（1,800件強）

公開不可または公開困難 47%（1,800件弱）

（公開不可の理由～ 条例等、内部情報、個人情報）

収集したデータの活用と可能なもののオープン化

庁内利用：簡易データベースとしてH27.6から庁内公開（全3,600件）

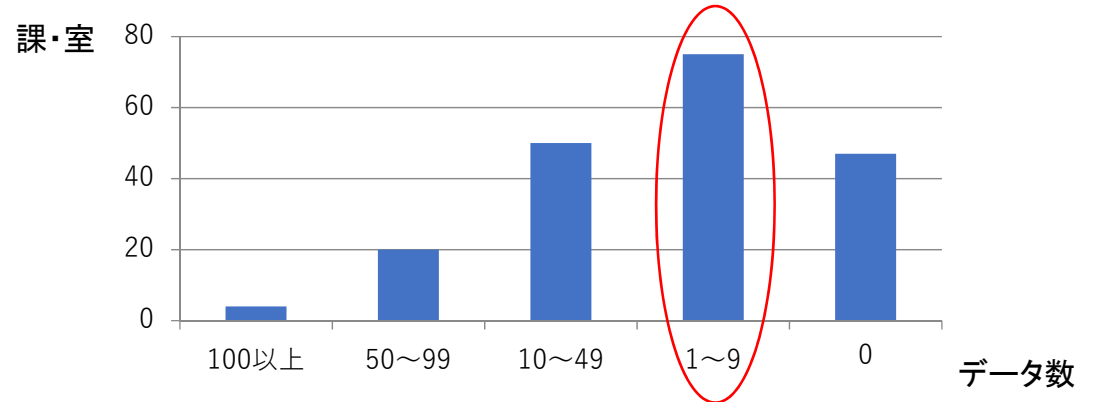
オープン化：公開可能または一部公開可能なものの中で、データの公開が容易な形式のものを各課から収集しH27年度末に公開予定

3.5 オープンデータ～実際の取組で見えてきた課題(2)

全庁から確実にデータを集めることは結構難しい

1桁回答が75の課室

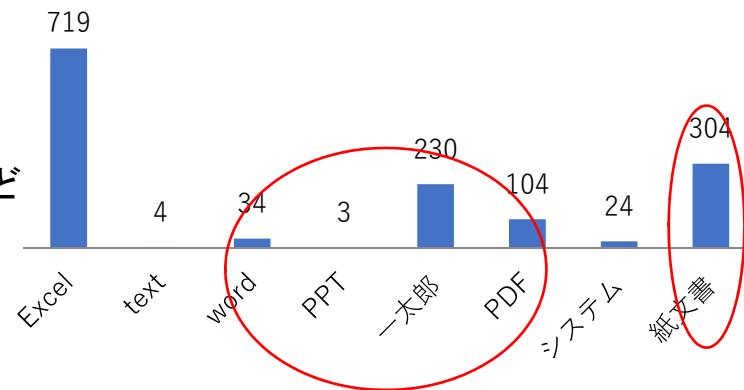
- ・所有しているデータの一部しか記載していない所属が多い



データ管理のリテラシーに課題

47%超が望ましくない所有形態

- ・「紙」で保管している
- ・表形式のデータもワープロ等で作成している など



3.6 オープンデータの推進の基本的考え方

大局的視点

「デジタル化された様々なデータを有効に活用(分析)することで地域振興に寄与すること」

「オープン化」と「利活用推進」を
総合的に推進することが重要！

【取組①】

庁内所有データのオープン化

- ・庁内データの棚卸
- ・公開サイトの準備
- ・データの収集とクレンジング
- ・データ公開(計画的かつ段階的に)

- ・データ活用推進は、「オープンデータ」に限らず様々なデータを想定すること。『IoT, BigData, AI』もデータ活用に関すること
- ・企業が独自に集めるプライベートデータにオープンデータをどう組み合わせるかが重要

【取組②】

地域におけるデータ活用推進

- ・EBPMなど庁内での率先的な活用！
- ・地域企業等の意識改革
(庁内向け勉強会、地域向けセミナー等)

【目的】

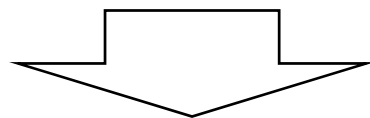
地域産業振興・新たな公共

- ・組織横断的な政策立案、横連携強化
- ・地域における事例の創出
(売上拡大、生産性向上、雇用拡大等)

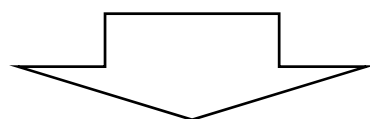
4. データ活用のポイント

4.1 なぜデータ分析を行うのか？

「データ分析」は手段であり目的ではない
何のためにデータ分析を行うのか？




「課題」解決のためにデータ分析を行う！

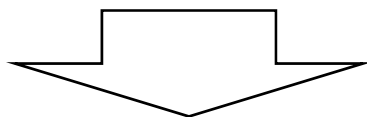


そもそも「課題」とは何なのか？

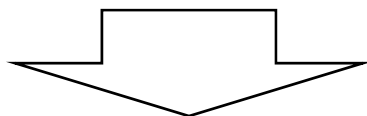
4.2 課題とは？

あなたの組織（部署）の課題は何ですか？

 ※この問いに明確に答えられないケースが多い



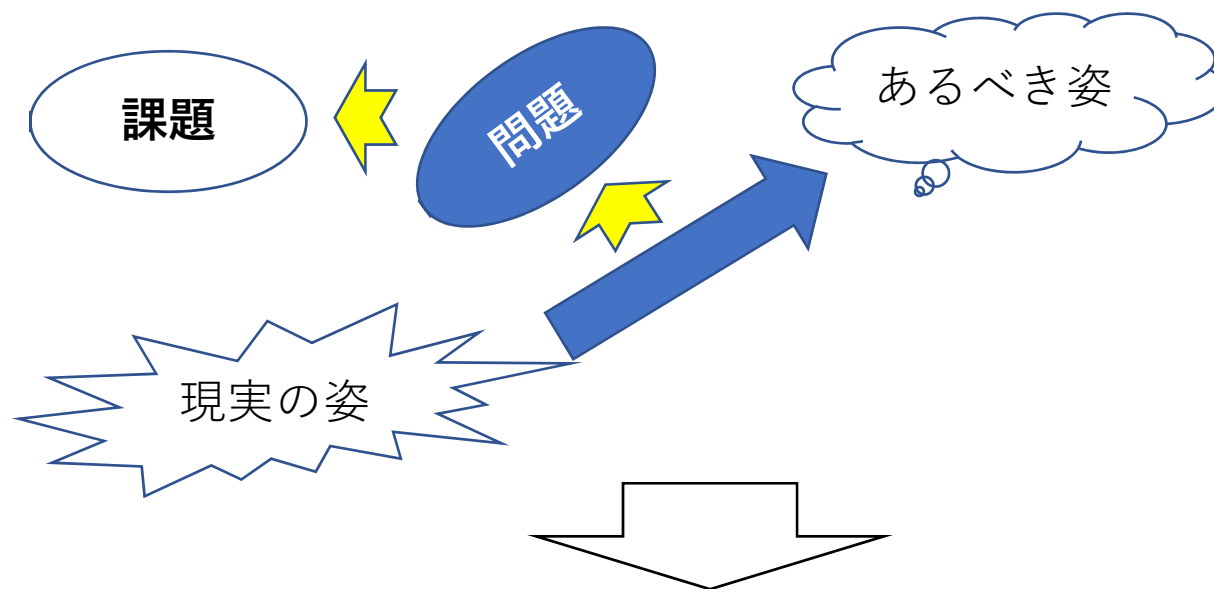
「課題」とは「問題」を解決するための行うべき事項



それでは「問題」とは何なのか？

4.2 課題とは？(2)

「現実」と「あるべき姿」のギャップ＝「問題」



まずは、「現実の姿」「あるべき姿」を
組織（部署）の関係者で共通認識することから！

4.3 課題を共通認識するために

現実の姿

あるべき姿



・観光客が少ない



・観光客を増やす

【現状を客観的数字で整理】

- ・年間の外国人観光客が〇人である
- ・過去5年間の変動は〇表のとおりである
- ・宿泊外国人数は〇人である
- ・隣県は〇人と〇倍の観光客数である
- ・その他同規模、類似の県の観光客数は〇人である

【ゴール（目標）の設定】



・観光客を〇人に増やす

そのためには・・・

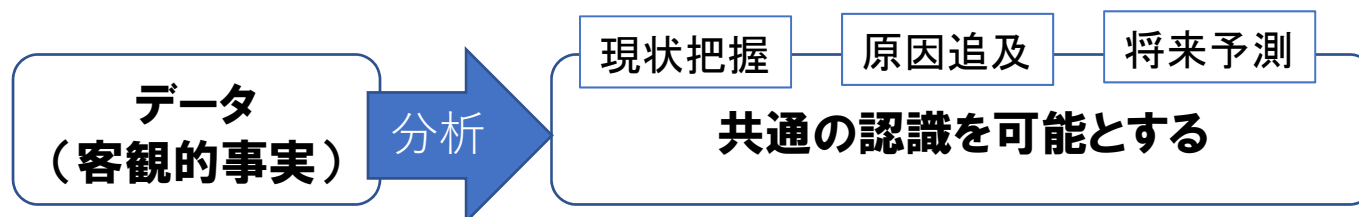
KGI

KPI

【ゴールに向かうための具体的な中間目標の設定】

4.4 データを使った客観的な評価・予測

統計による分析



【統計】に関して

細かな計算式を覚える必要はないが（ツールに任せる）
大まかでもいいので意味を押さえておくと良い

4.5 現実を正しく把握すること(1)

	A課	B課
4月	25.23	12.02
5月	1.50	8.38
6月	8.52	6.68
7月	7.50	5.79
8月	9.35	4.32
9月	17.25	5.00
10月	6.00	7.50
11月	3.25	7.91
12月	18.24	13.75
1月	3.75	11.86
2月	15.75	12.60
3月	2.25	17.99
年合計	118.59	113.80
月平均	9.9	9.5

「合計」「平均」だけを見ると差がないが
「バラつき」に差がある

偏差：データから平均を引いた値
平均からの距離（プラス、マイナス）
総和は0になる

分散：偏差の2乗の総和の平均
バラつきの大きさ

標準偏差：分散の平方根
バラつきの大きさ（単位を戻す）

A課の標準偏差：7.23

B課の標準偏差：3.96

(参考) RPA (Robotic Process Automation)

働き方改革に有効な
新たなツール

RPAとは

これまで人間が行ってきた定型的なパソコン操作を
ソフトウェアのロボットにより自動化するもの

総務省ホームページより http://www.soumu.go.jp/menu_news/s-news/02tsushin02_04000043.html

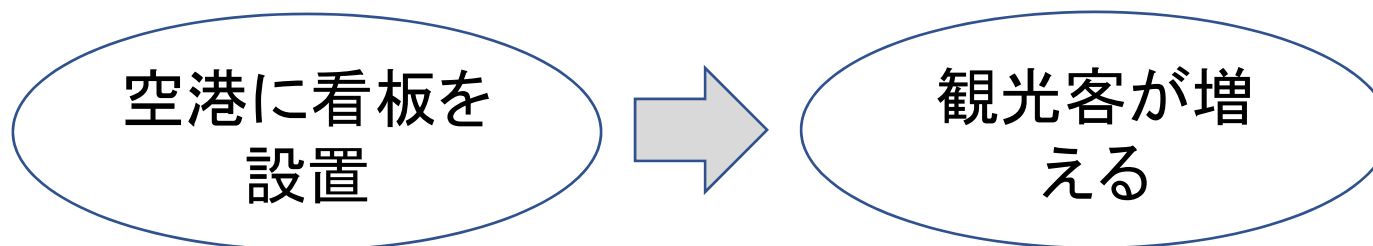
■今後の展望

RPAは、これまで人間のみが対応可能と想定されていた作業、もしくはより高度な作業を人間に代替して実施できるルールエンジンやAI、機械学習等を含む認知技術を活用した業務を代行するツールになりつつあります。人間の補完として業務を遂行することから、仮想的労働者(Digital Labor)として、2025年までに事務的業務の1/3の仕事がRPAに置き換わるインパクトがあるともいわれています⁹。

RPAの導入と運用は、働き方改革につながる「業務改善・改革」の目標や方向性を明確にし、自社に合うツール選択とマネジメント方法に配慮して進めることが重要といえるでしょう。実装による自動化に伴う変化を見極め、現場部門とシステム関連部門が連携し、運用局面ごとのルール(シナリオ)策定や適用可能な業務領域の拡大を検討するなど、継続的にPDCAサイクル¹⁰をまわしながら活用していく姿勢が望まれます。

4.6 相関

相関：2つのものごとの変化に関連性があること



本当に効果があるのか？

看板の数を増やすほど観光客が増える？！

<注意> 相関関係があったとしても因果関係にあるとは言えない！

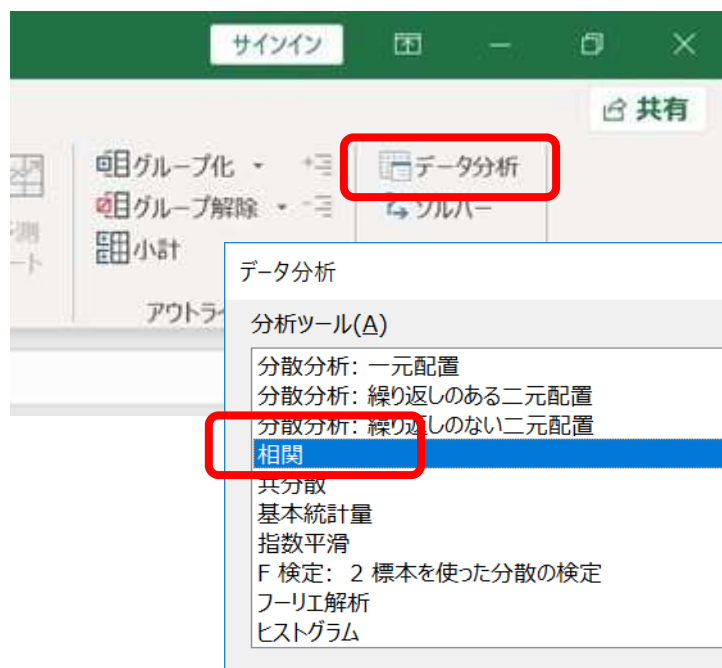
(例えば)

観光客が増えたのは、たまたま看板設置の翌週から
大きな祭りの開催が重なったため

4.7 相関係数

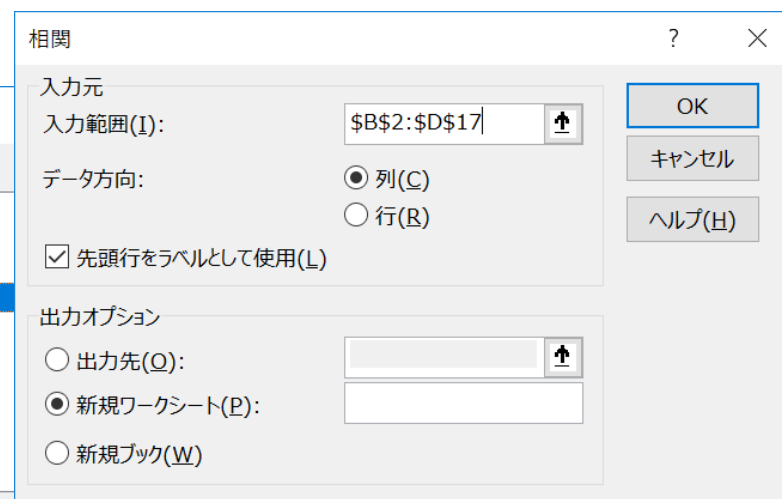
相関係数 = 2種類のデータの（直線的な）関係性の強さを
- 1 から 1 までの数値で表したものの。

Excelのデータ分析ツールで
容易に計算できる



相関係数の目安

0.7 ~ 1.0	かなり強い正の相関がある
0.4 ~ 0.7	正の相関がある
0.2 ~ 0.4	弱い正の相関がある
-2 ~ 0 ~ 0.2	ほとんど相関がない
-4 ~ -2	弱い負の相関がある
-7 ~ -4	負の相関がある
-10 ~ -7	かなり強い負の相関がある

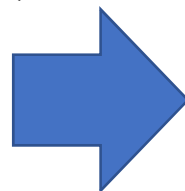


4.7 相関係数(2)

ある月の売上個数

	ビール	かき氷	だんご
1日	11	19	45
2日	17	20	30
3日	23	11	96
4日	25	26	60
5日	30	28	69
6日	31	47	49
7日	43	40	25
8日	47	35	53
9日	62	47	78
10日	61	52	83
11日	75	39	66
12日	80	87	47
13日	82	85	77
14日	85	78	55
15日	100	85	77

Excel
ツール



相関係数

	ビール	かき氷	だんご
ビール	1		
かき氷	0.8969	1	
だんご	0.294485	0.108326	1

ビールとかき氷：0.89（強い相関）

ビールとだんご：0.29（弱い相関）

かき氷とだんご：0.11（相関なし）



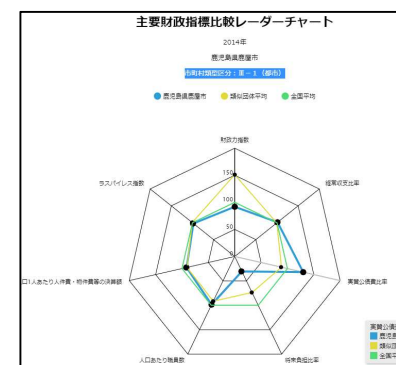
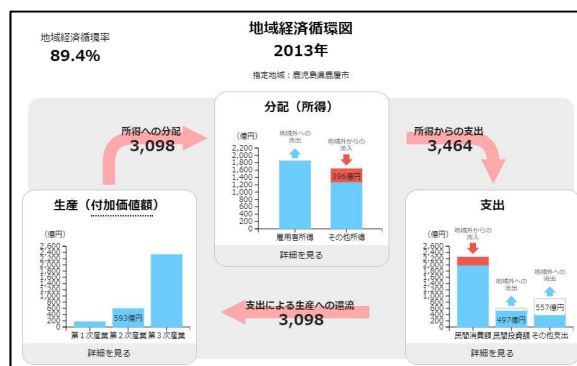
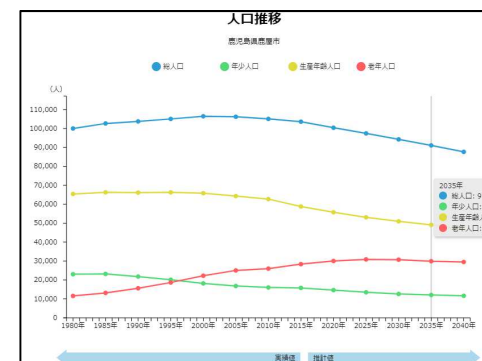
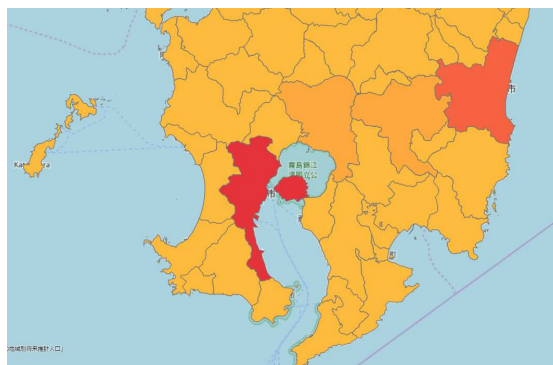
**ただし、ビールを一生懸命売っても
かき氷の売上を伸ばすことには繋がらない
（相関関係にあるが因果関係はない）**

擬似相関（ぎじそうかん、英: Spurious relationship, Spurious correlation）は、2つの事象に因果関係がないのに、見えない要因（潜伏変数）によって因果関係があるかのように推測されること。



4.8 RESAS(地域経済分析システム)

RESASは、日本・内閣府のまち・ひと・しごと創生本部が運用している、産業構造や人口動態、人の流れなどに関する官民のいわゆるビッグデータを集約し、**可視化**を試みるシステム



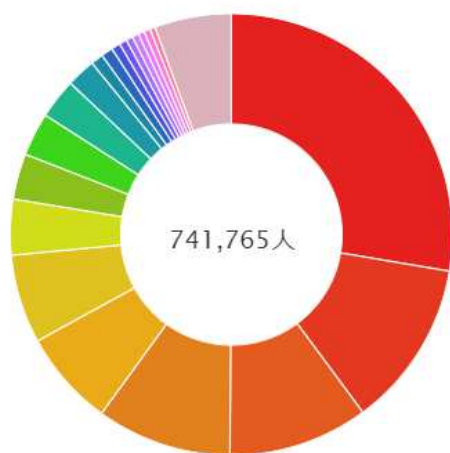
4.8 RESAS (2) (外国人観光客の入出空港組合せ)

広島県

入国空港 x 出国空港内訳

表示年・期間：2016年すべての期間

指定地域：広島県



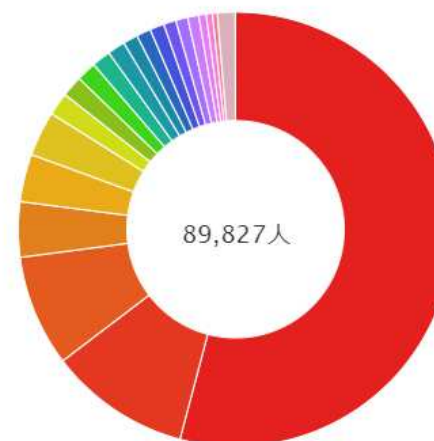
- 1位 成田国際空港 - 成田国際空港 205,165人 (27.66%)
- 2位 関西国際空港 - 関西国際空港 90,712人 (12.23%)
- 3位 東京国際空港 (羽田空港) - 東京国際空港 (羽田空港) 75,861人 (10.23%)
- 4位 広島空港 - 広島空港 73,321人 (9.88%)
- 5位 成田国際空港 - 関西国際空港 52,091人 (7.02%)
- 6位 福岡空港 - 福岡空港 48,097人 (6.48%)
- 7位 東京国際空港 (羽田空港) - 成田国際空港 30,217人 (4.07%)
- 8位 関西国際空港 - 成田国際空港 24,804人 (3.34%)
- 9位 東京国際空港 (羽田空港) - 関西国際空港 23,471人 (3.16%)
- 10位 成田国際空港 - 東京国際空港 (羽田空港) 21,771人 (2.92%)

山口県

入国空港 x 出国空港内訳

表示年・期間：2016年すべての期間

指定地域：山口県



- 1位 福岡空港 - 福岡空港 48,618人 (54.12%)
- 2位 成田国際空港 - 成田国際空港 9,450人 (10.52%)
- 3位 関西国際空港 - 関西国際空港 7,456人 (8.30%)
- 4位 東京国際空港 (羽田空港) - 東京国際空港 (羽田空港) 3,676人 (4.09%)
- 5位 広島空港 - 広島空港 3,199人 (3.56%)
- 6位 その他 - 福岡空港 2,993人 (3.33%)
- 7位 東京国際空港 (羽田空港) - 成田国際空港 1,545人 (1.72%)
- 8位 中部国際空港 - 福岡空港 1,404人 (1.56%)
- 9位 その他 - 成田国際空港 1,336人 (1.49%)
- 10位 成田国際空港 - 東京国際空港 (羽田空港) 1,256人 (1.40%)

4.9 データ分析ツール

Excel

単一データをもとにした表やグラフの作成、簡易な統計分析
(基本的な分析には十分使える)

統計解析ツール

SPSS (IBM)

JMP (SAS Institute)

R (フリーソフト)

.....

BIツール

tablaue (tablaue japan)

Oracle BI (オラクル)

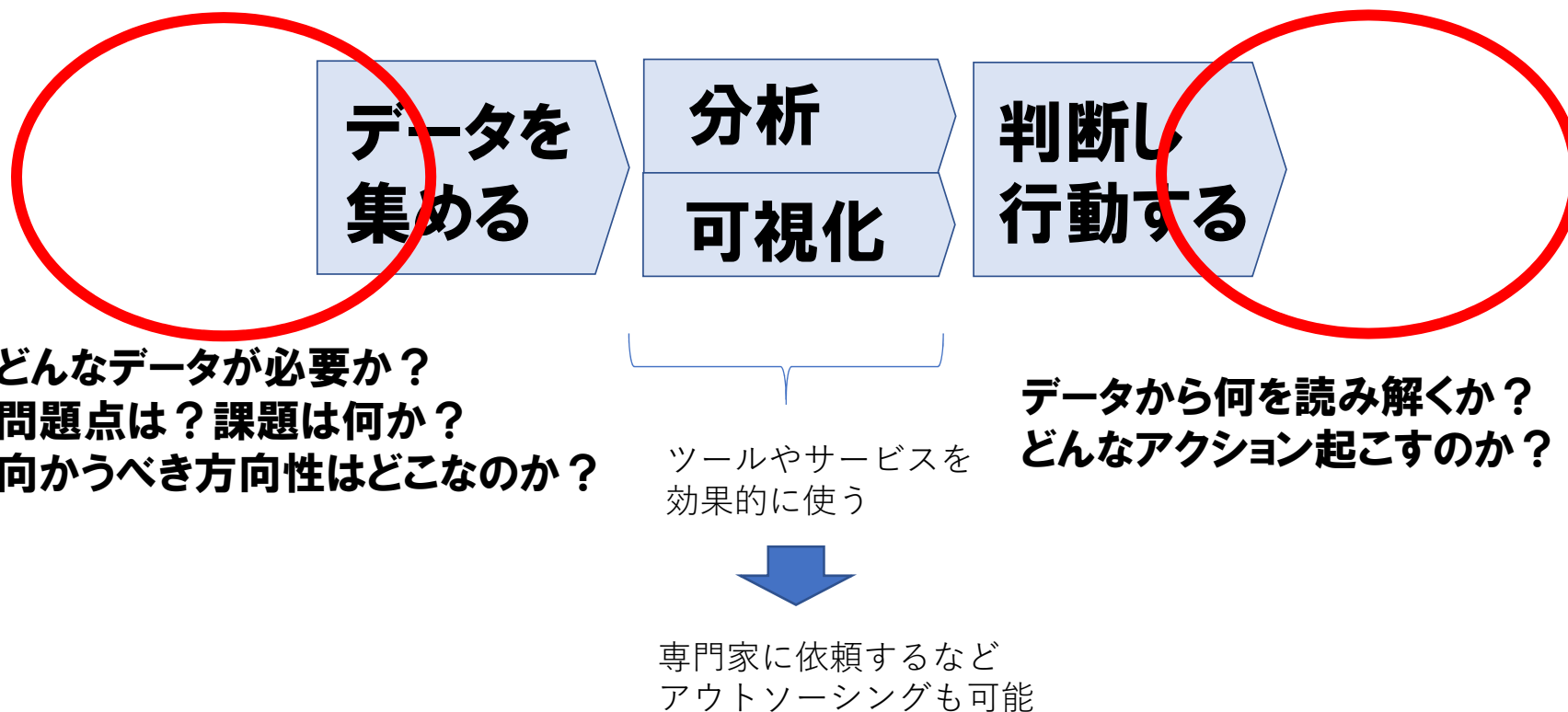
.....



大容量、多数のデータ、
複数人での作業共有、
ビジュアル性の高いレポート等
業務によっては導入を検討！！

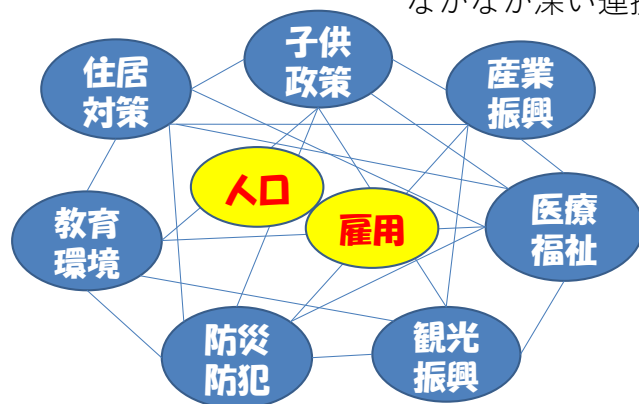
4.10 力点をどこに置くか

最初と最後は特に重要！

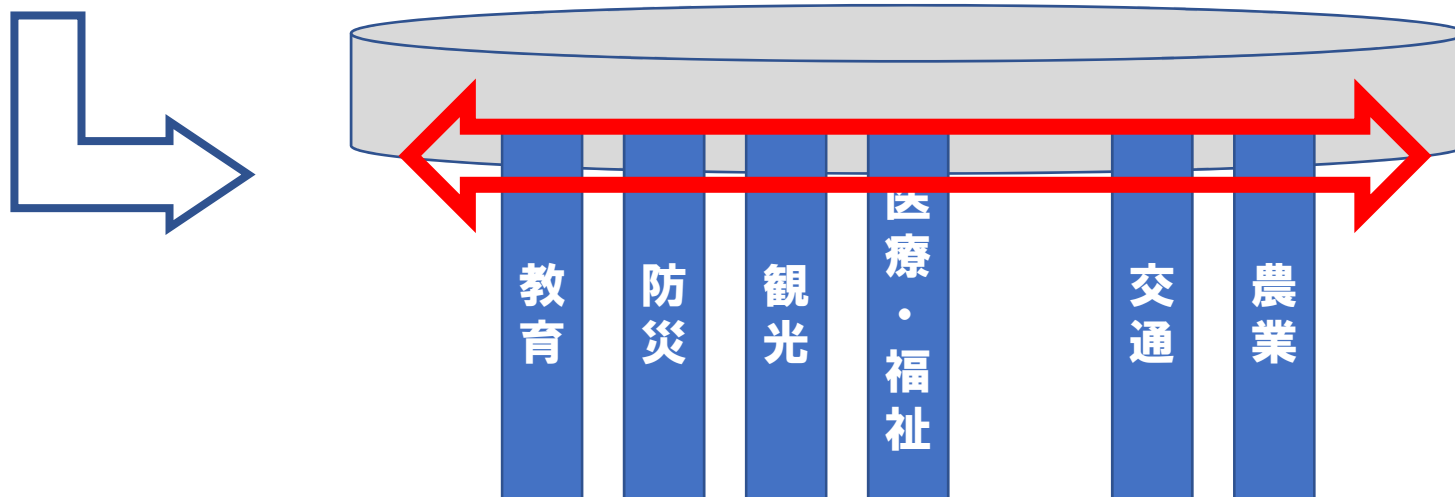


4.11 データ活用時代は、本格的な横連携のはじまり

現実の業務でも様々な分野が関係することは多いが
なかなか深い連携には至っていない



**各部署が保有するデジタルデータを
共有することから
縦割りを超えた横連携のスタート！**



4.12 さいごに

- ・庁内保有データの100%デジタル化と保存ルールの整備
- ・庁内の情報連携(データの相互活用)環境の整備



ICT利用からデータ活用への転換

使えるデータを探すのは結構大変

内部で分析事例なども共有していき、ノウハウを伝承・伝搬していく

- ・EBPMの推進など行政自身のデータ活用の高度化



行革担当や企画担当との連携強化

- ・データを見る目(判断する能力)を鍛えること