

情報科学 【AI・データサイエンス】

第4回 データの収集

データ収集とバイアス
適切に守られるべきデータと積極的に活用すべきデータ

データ収集とバイアス

調査のためのデータ収集

データをどうやって集めるか？

調査とデータ収集

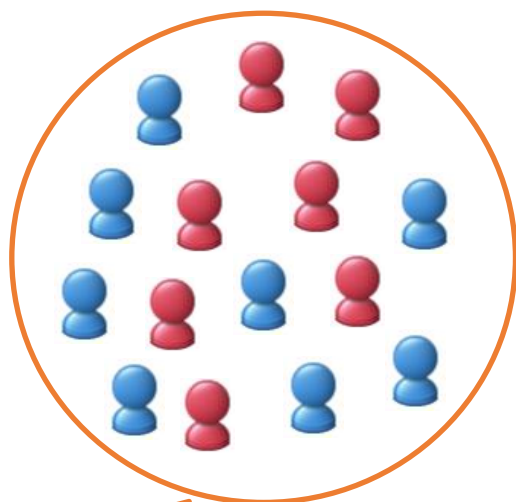
- 統計調査を使って何かを調べたい
 - 内閣支持率を調査したい
 - 日本の女性の平均身長を調べたい
 - 労働者の平均年収を調べたい
 - ...
- できる限り正確なデータに基づいて調査したい
- さあ、どうやってデータを集めるか？



理想的状況：全数調査

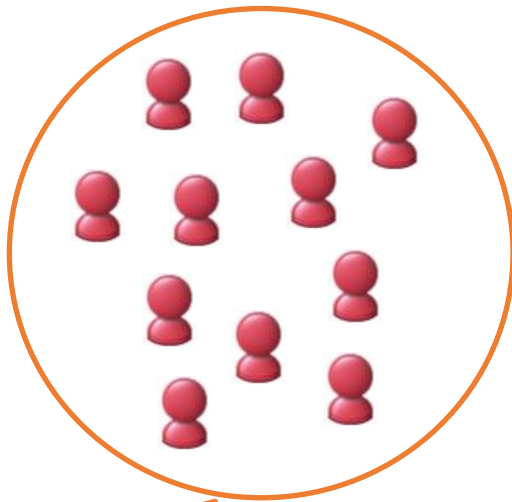
- 調査対象をすべて調べること
 - すべての調査対象は「母集団※」と呼ばれる

例) 内閣支持率



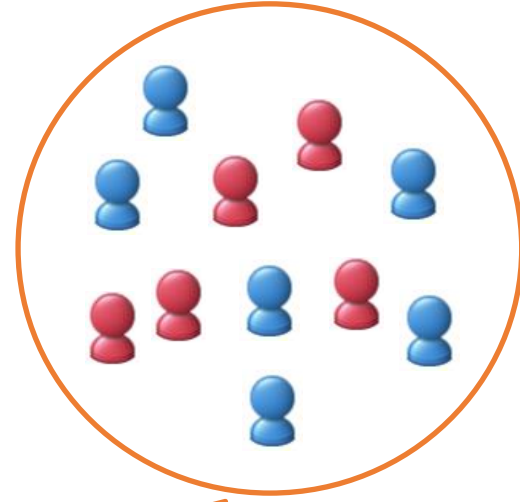
国民全員を調査

例) 日本人女性の平均身長



日本人女性全員を調査

例) 労働者の平均年収



労働者全員を調査

※「おおもとの集団」という意味で、お母さんという意味はない。英語ではpopulation. ちなみに母集団はあっても父集団はない。

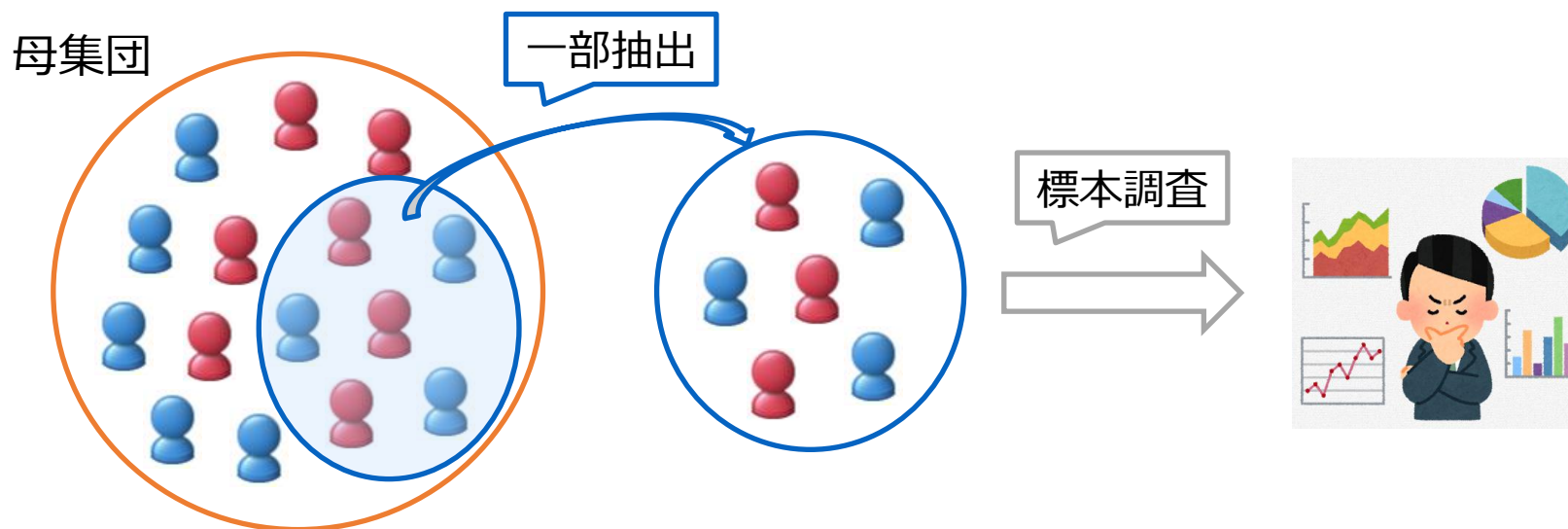
全数調査

- 全数調査が使える場面
 - すべての母集団を調査できる環境が整っている (例: 国勢調査)
 - 母集団の規模が小さいとき (例: クラス内の学生の成績調査)
- 全数調査のメリット
 - 調査で得られたデータは**真実**
- 全数調査のデメリット
 - 一般的な統計調査で全数調査は事実上**不可能**
 - 例: アンケート調査, 工場での部品の品質検査
 - 調査に対する労力が単純に大きい



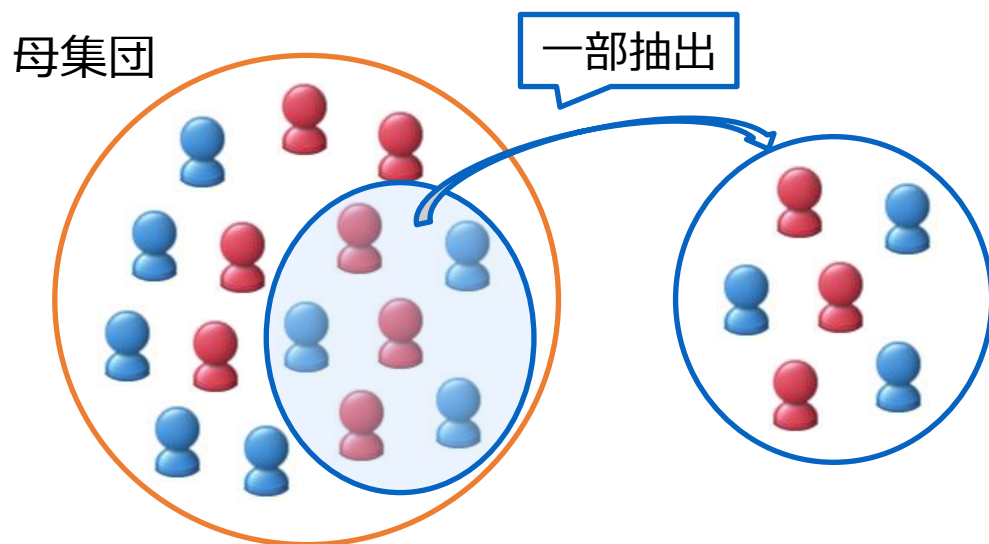
全数調査が無理なら「標本調査」

- 調査対象となる母集団の一部を取り出して調査すること



標本調査

- 調査対象となる母集団の一部を取り出して調査すること



標本の抽出には2種類ある
(標本抽出法)

- 有意抽出法
- 無作為抽出法

有意抽出法

- 調査の企画者(ら)が独自の判断(ある一定の法則や何かしらの固定観念)で標本を抽出
- 有意抽出法の代表例: 便宜抽出
 - 収集しやすい調査対象から標本を抽出すること

有意抽出によるアンケートの例①

調査内容: 若者の流行りのファッション
調査方法: 街頭アンケート
調査場所: 繁華街
調査時間帯: 13時～17時



問題

- その時間帯に出歩けない人もいる
- 繁華街を普段歩かない人もいる
- アンケートを拒否する人もいる

有意抽出によるアンケートの例②

調査内容: 政党支持率調査
調査方法: 固定電話調査
調査時間帯: 13時～17時



問題

- その時間帯に電話に出れない人もいる
- 固定電話を持っていない人もいる
- アンケートを拒否する人もいる

有意抽出法のメリット・デメリット

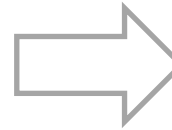
- メリット

- 調査対象を「独自の判断で」絞り込めるので、データ収集が楽

調査企画者の独自判断

- デメリット

集めやすいから
こういう人に聞こう



- 独自判断で調査が行われるため、調査結果が母集団全体を代表していないことが多い
 - 要するに、手元のデータがすべてを表していない！ → 標本選択 **バイアス**（次スライド）
- 独自判断が誤っていれば、真実とは異なる結果が出てくる可能性

参考

http://rplus.wb-nahce.info/statsemi_basic/musakuityusyututoyuityusyutu.html

<https://www.stat.go.jp/teacher/c2hyohon.html>

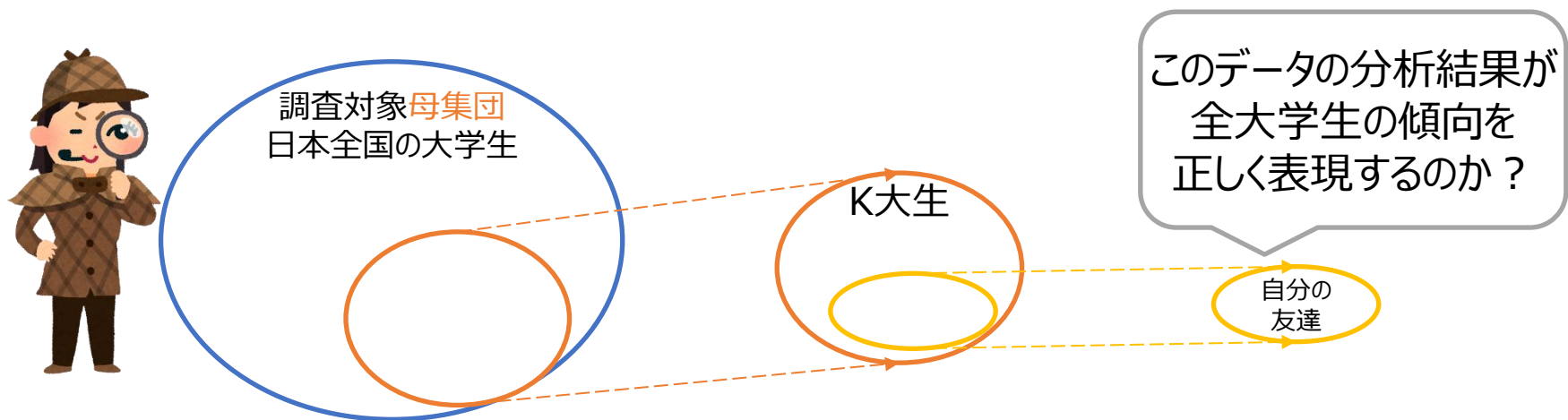
標本選択バイアス

あなたの集めたデータ、偏ってませんか？



標本選択バイアス

- バイアス＝「偏り」のこと
- 「全大学生の勉強時間と試験の点数の関係を調べたい」
 - それを自分の友達だけで調べたら...



なるべく標本選択バイアスを入らないように データを収集するには

● 母集団から直接ランダムにデータを選ぶ

● 単純無作為抽出法

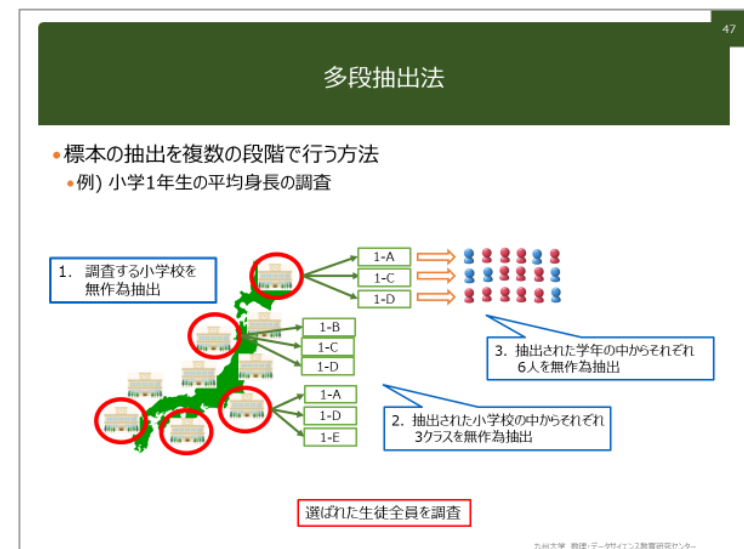
- 母集団が6人なら, サイコロを振って出た目の番号の人のデータを収集
- 母集団が1000人なら, 1～1000の目のあるサイコロを振って. . .

ランダム =



● 他にもいろいろ(→付録「バイアスの少ない標本調査法」)

- クラスター抽出法 (集落抽出法)
- 二段抽出法
- 多段抽出法
- 層化抽出法



その他, ありがちな標本選択バイアス

- 既已取得されたデータ (= 過去のデータ) から未来を予測
 - 過去の状況と未来の状況が同じという保証はない
- 1年間ジムに通い続けた人の, 減量効果
 - 途中でジムをやめてしまった人については分析対象外
 - もしも, 効果があった人だけが1年間通い続けたとすれば, 減量効果が過大評価されてしまう可能性がある
 - 生存者バイアスとも呼ばれる
- 現社員全体のアンケートに基づく, 方針決定
 - 年齢や性別におけるマイノリティが差別される可能性も



似たような話：「メディアの街頭インタビュー」

- 恣意的に選んだ意見のみを，あたかも一般論のように報道するなら，最も悪質な「標本選択バイアス」

そもそも放送される「わずか2,3人」の意見が世の中の代表である保証なんてまったくない

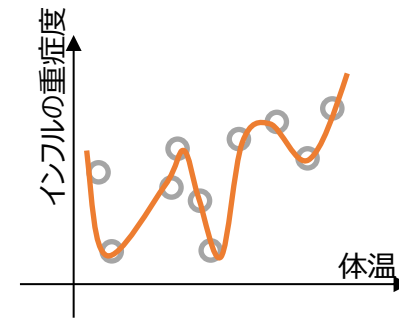
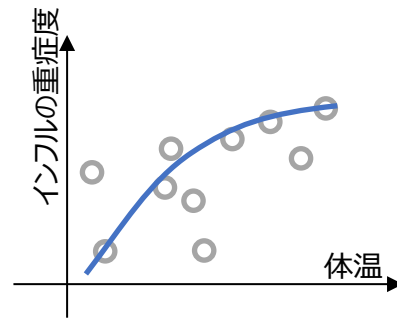
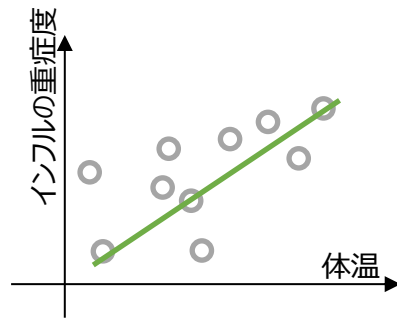
- 皆さんもだまされないように気を付けましょう
- 結構，「あの手の手」でやってきますよ



谷岡，「社会調査」のウソ

参考：標本選択バイアス以外のバイアス： 帰納バイアス

- データ分析の方法には様々なものがある
 - 同じデータであっても，方法によって「予測」結果はいろいろ



どの方法を使うかは
私の判断次第



- 帰納バイアス = 利用した方法の性質に依存したバイアス
 - 「この方法だからこそ，この結果になった」ということ
 - = 「違う方法だと違う結果になってしまう」
- これも避けるのはなかなか難しい！
 - 選択した方法の妥当性を適切に評価する必要がある
 - 参考→バリデーション，AICやBICなどの(情報量)基準

参考：標本選択バイアス以外のバイアス： アノテーションバイアス

- **アノテーション** = データの**説明**のこと
 - 例えば「それが何の画像であるか」を説明
 - **だれか人間が**、この説明を付与する
 - 画像認識AIは、このアノテーションを教師として学習される



- ところで、そのアノテーション、**本当に信じられるの？**
 - 所詮は人間がつけたもの。先入観も入り得る
 - それでAIを学習すると...



適切に守られるべきデータと 積極的に活用すべきデータ

個人情報とオープンデータ

個人情報とデータ



個人情報保護法の概要

個人情報保護法

- 正式名称「個人情報の保護に関する法律」
- その名の通り、個人情報の保護を目的とする法律(第1条)
 - ただ単に情報の利用を制限するだけではない
 - 個人情報の適正かつ効果的な利用が、産業の創出等に有用であることを踏まえた上で、個人の権利利益を守ることを目的とする → 両者のバランスが重要



- 経緯は→付録

個人情報保護法による「個人情報」の定義

- 生存する個人に関する情報であって、以下のいずれかに該当するもの
 - 氏名、生年月日など、特定の個人を識別することができるもの
 - 他の情報と容易に照合することができ、それにより特定個人の識別が可能になるものを含む
 - 個人識別符号が含まれるもの
 - 身体の特徴を電子化したもの(生体情報)
 - 例: DNA, 容貌, 虹彩, 声紋, 指紋など
 - 個人がサービスを利用 or 商品を購入する際に割り当てられる符号
 - 例: 旅券番号, 免許証番号, 住民票コード, 個人番号



個人情報データの分析は可能か？ 個人情報の第三者提供

- 原則として、**本人への同意なしには無理**
 - 本人＝個人情報によって識別される特定の個人
- 例外①人の生命・身体・財産の保護のために必要がある場合
 - 例：緊急の事態にプロバイダが発信者情報を開示
- 例外②つぎの条件をすべて満たすとき
 - 本人の意向により「**提供停止（オプトアウト）**」可能な場合
 - 下記を本人に知らせておく
 - 第三者への提供を利用目的とすること、そのデータ項目、提供の方法
 - オプトアウトできること、さらにオプトアウトの受付方法
 - 個人情報保護委員会に届け出
- 例外③匿名加工情報になっている場合（後述）

簡単に言えば
「私のデータは
公開しないで！」と
除外してもらえる

ただし、
要配慮個人情報
(次スライド)は
オプトアウト可能でも
第三者提供NG

要配慮個人情報

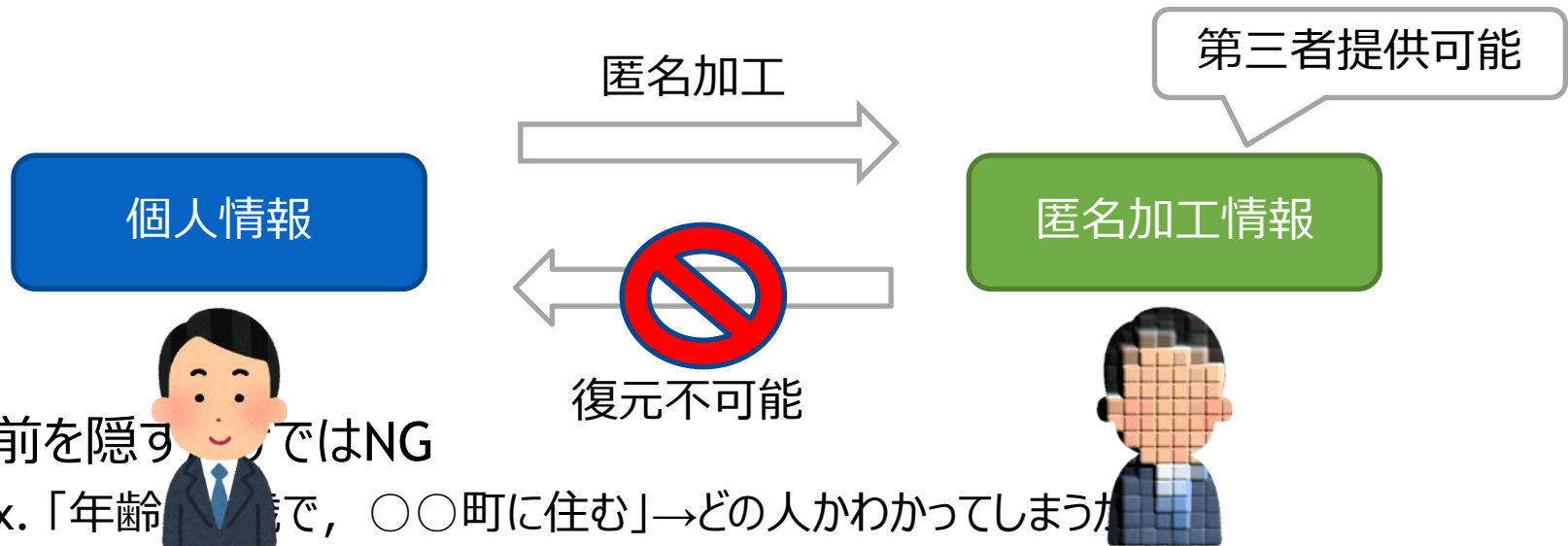
- 本人の人種，信条，社会的身分，病歴，犯罪歴，犯罪被害の事実，及び本人に対する不当な差別偏見が生じる恐れがあると政令で定める要配慮情報



- 要配慮個人情報は「本人の同意を得ない取得の原則禁止」
 - 取り扱いが特に厳重

匿名加工情報

- 個人を特定できないように加工した個人情報
 - 匿名加工情報の第三者提供する旨の明示・公表等は必要



オープンデータ

誰もが使える公共データを活用して、より良い世の中へ！



オープンデータの定義

- 行政機関(国や地方自治体)が保有する**公共データ**(※)のうち
 - ① 二次利用可能で
 - ② 機械判読に適している
 - ③ 無償で利用できる, 公開データのこと
- 要するに, 「市民や企業が使いやすい公共データ」
 - 二次利用可能 = そのデータを営利 (お金儲け) 目的に使ってもOK
 - 機械判読 = PDFや画像ではない「数値」データ
 - 二次利用・機械判読の詳細→付録



※ 政府や独立行政法人・自治体だけでなく, 公益企業など民間事業者や個人が保有し, 二次利用可能な形で公開されるものも含む

オープンデータの意義・目的： 政府の「オープンデータ基本指針」より

- 国民参加・官民協働の推進による諸課題の解決，経済の活性化
 - 行政の高度化・効率化
 - 透明性・信頼性の向上
 - 地域の課題の解決
-
- 政策は，「声の大きな人」の意見で決まるのではなく，データに基づいた客観的な分析に基づいて決まるべき
 - Evidence-based policy making (EBPM)と呼ばれる



行政（国や自治体）によるオープンデータ推進



日本政府
(国)

ぼくらが管理してる公共のデータを、
・2次利用OKで、
・みんなが使いやすい形式で、
・かつタダで使えるようにしたら、
新事業がバンバン生まれるかも！？
特にスマホとかSNSとか、
今すごく普及してるし！

新しい事業が
地方で生まれたら、
地域の活性化・振興に
役立つね！
ぼくらの仕事（政治）の
透明性や市民の理解も
深まるかも！



地方自治体
(市町村)

オープンデータを推進しよう！

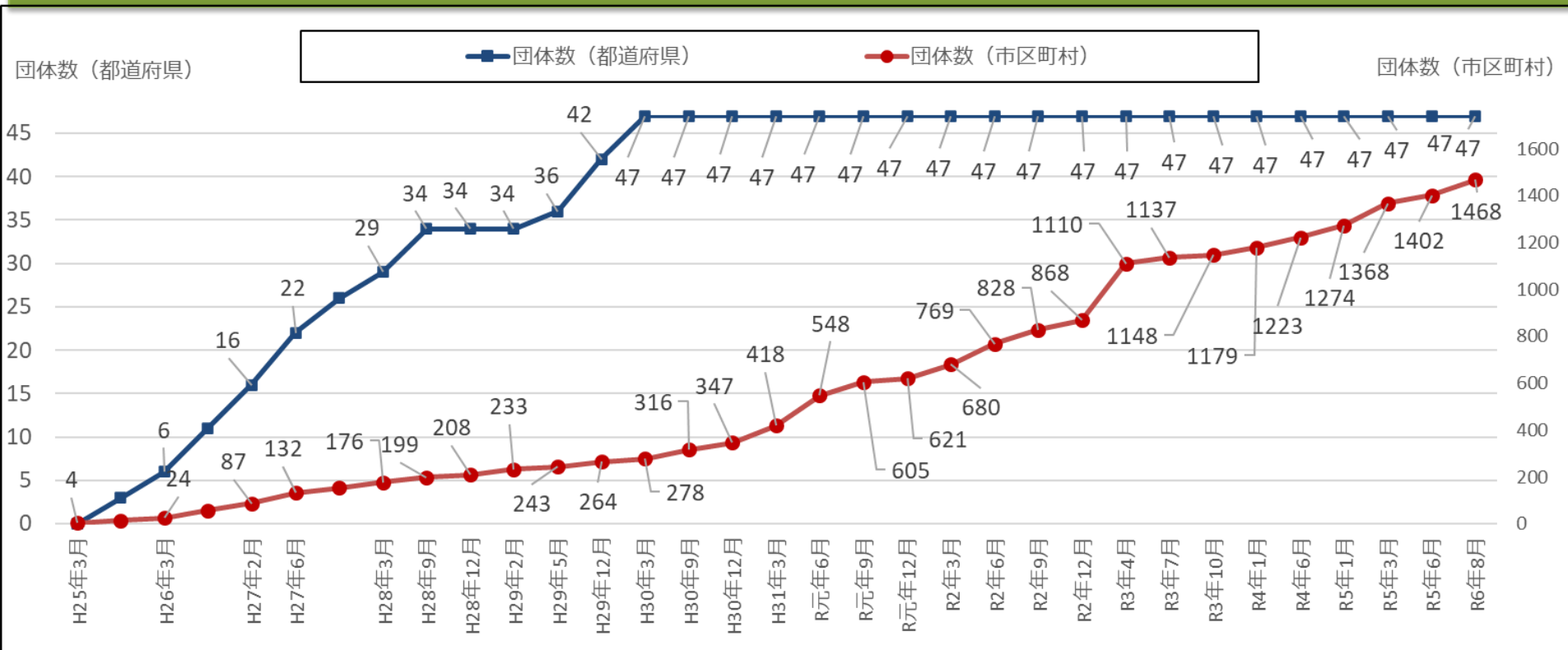
官民データ活用推進基本法（平成28年法律第103号）第11項の規定により、地方自治体に対し、自らが保有する官民データについて国民が容易に利用できるよう必要な措置を講じることを義務化

さらに

平成29年に閣議決定された「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、令和2年度までに地方自治体のオープンデータ取組率100%を目標とすることが明記

オープンデータに関する地方自治体の取り組み状況 →デジタル庁HPで確認可能

地方公共団体のオープンデータ取組済み（※）数の推移



https://www.digital.go.jp/resources/data_local_governments

オープンデータと著作権

クリエイティブ・コモンズ



「ホームページで公開＝オープンデータ」ではない (1)

たとえば、福岡市のホームページ！

福岡市 FUKUOKA CITY

本文へジャンプ よくある質問(Q&A) English 中文 韓国語

ホーム スマホサイトへ 音声読み上げ 文字サイズ 小 中 大

Google カスタム検索

サイト内検索について

防災情報 救急医療・消防

暮らし・手続き 子育て・教育 健康・医療・福祉 観光・魅力・イベント 創業・産業・ビジネス 市政全般

暮らし・手続き

ページ内メニュー

- 届出・証明・税金
- ネットで手続き

暮らし・手続きの新作情報

福岡市出前講座「7-6 ペットボトルのリサイクル(当公民館)」を実施しました。 2019年9月26日 NEW

Google カスタム検索

サイト内検索について

防災情報 救急医療・消防

暮らし・手続き 子育て・教育 健康・医療・福祉 観光・魅力・イベント 創業・産業・ビジネス 市政全般

3 ダウンロードできる様式など

市民税・県民税申告書

平成31年度市民税・県民税申告書 (1,306kbyte)

平成31年度市民税・県民税申告書の書き方 (5,261kbyte)

市民税等に関する資料がPDFで公開されています。

Google カスタム検索

暮らし・手続き

ホームページで公開されている資料等が著作物に当たる場合、著作権法により権利が保護されていることとなります。

著作権について

福岡市ホームページに掲載している個々の情報(文章、写真、イラストなど)に関する著作権は、原則として福岡市に帰属します。(一部の画像等の著作権は、福岡市以外の原作者が所有しています。)

当ホームページの内容について、「私的使用のための複製」や「引用」など著作権法上認められた場合を除き、無断で複製・転載することはできません。

なお、許可されたルールの範囲内で、商用利用も含め自由に複製・加工や頒布などができるデータを、以下のサイトで公開しています。

福岡市オープンデータサイト (<https://www.open-governmentdata.org/fukuoka-city/>)

1頁 平成 31 年度 市民税 県民税 申告書 福岡市 区長

住所・氏名・職業等

住所 氏名 フリガナ 印

携帯電話 自宅電話 ()

住民票の住所 (住所と異なる場合記載)

職業又は勤務先 電話 () 屋号

生年月日 年 月 日生

世帯主の氏名 及び 続柄

種類	A 収入金額	B 必要経費	所得金額
1 給与等	円	円	円
2 不動産所得	円	円	円
3 雑所得	円	円	円
4 総合所得	円	円	円
5 所得金額の合計	円	円	円

所得金額の合計 ① ~ ⑤ + ⑥ =

「ホームページで公開＝オープンデータ」ではない（2）

- オープンデータとは...



- ① 二次利用可能で
- ② 機械判読に適していて
- ③ 無償で利用できる
公開データ！

- 一方、先ほどのPDF資料は...

- ① 2次利用可能？ → × 著作権で保護
- ② 機械判読に適している？ → × PDFは閲覧専用で分析不向き
- ③ 無償？ → ○

- つまり、ホームページに公開しただけではオープンデータではない

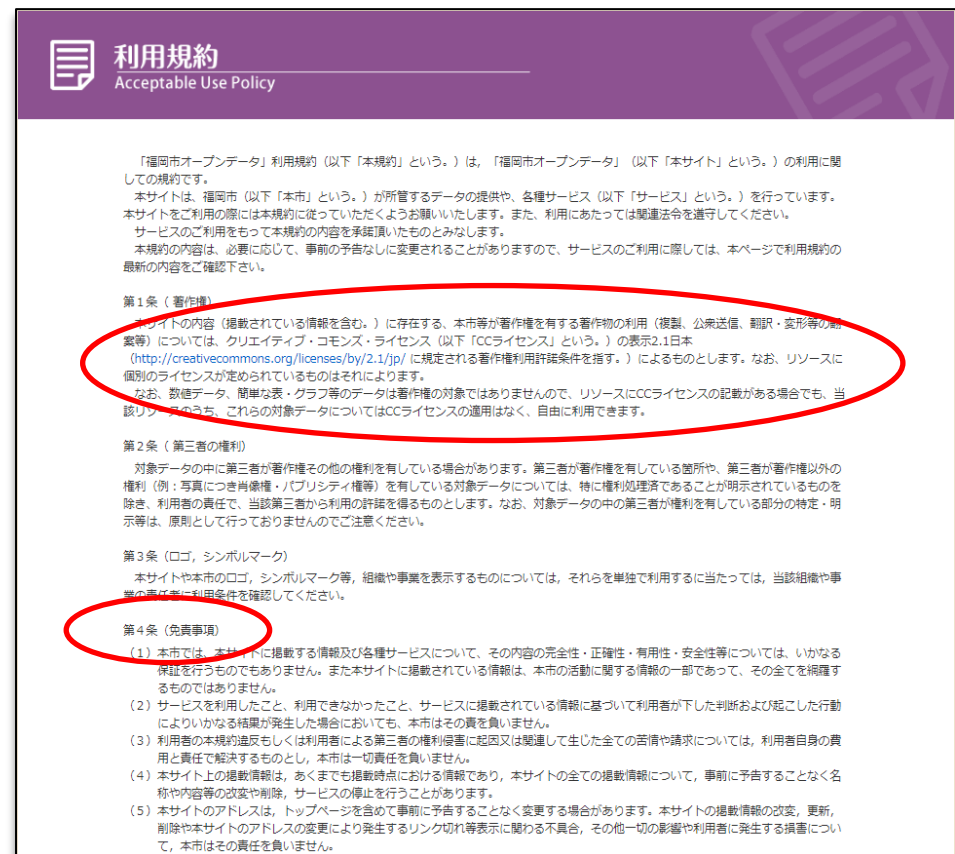
オープンデータと言うためには...

- ①二次利用のための条件を掲載し, ②機械判読に適した形式で, ③無償提供
- 例: PDFをCSVに加工して, クリエイティブ・コモンズ・ライセンスをつけて公開



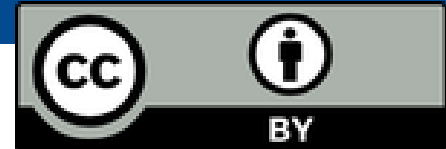
出典: 福岡市オープンデータ

(<https://www.open-governmentdata.org/fukuoka-city/>)





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
 - クリエイティブ・コモンズ・ライセンス（CCライセンス）を提供している国際非営利組織とそのプロジェクトの総称
- CCライセンス
 - インターネット時代のための新しい著作権ルール
 - 作品を公開する作者が「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
- CCライセンス利用で
 - 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可能
 - 受け手はライセンス条件の範囲内で再配布やリミックスなどが可能



クリエイティブ・コモンズ・ライセンスのレベル



全ての権利の主張



いくつかの権利の主張



全ての権利の放棄



BY : 著作権者の表示

Attribution



NC : 商用利用不可

Non-Commercial



ND : 改変不可

No Derivatives

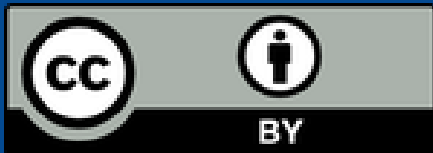


SA : 改変版についても,
レベルを継承

Share Alike

最も緩くてオープンデータ向き





を付けて公開することで

- 各市町のクレジット（市名、データ名等）を表示すれば，利用者が営利目的を含めて自由にデータを改変，複製，再配布することが可能に！



がないと..

- ・ホームページの情報を利用する際，著作権処理（使用許可等）に手間，時間，費用等がかかる
- ・利用に制約があり，利用者は自由に編集・加工ができない



があると..

- ・出典を明記すれば、利用者は、著作者の許可を得ずとも自由にホームページ情報の二次利用が可能
- ・利用者は、自由に編集・加工ができるため、他のデータとも組み合わせて利用拡大が見込める

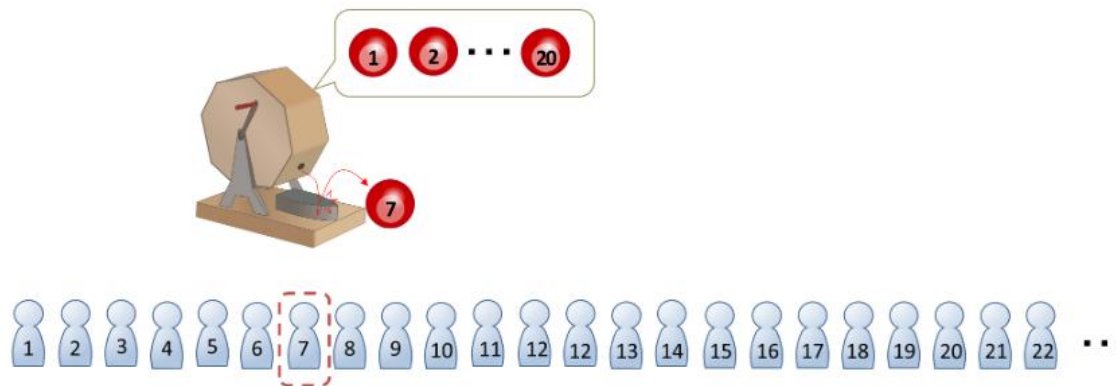
付録

バイアスの少ない標本調査法

主観の入らないデータをどうやって手に入れるか？

単純無作為抽出法

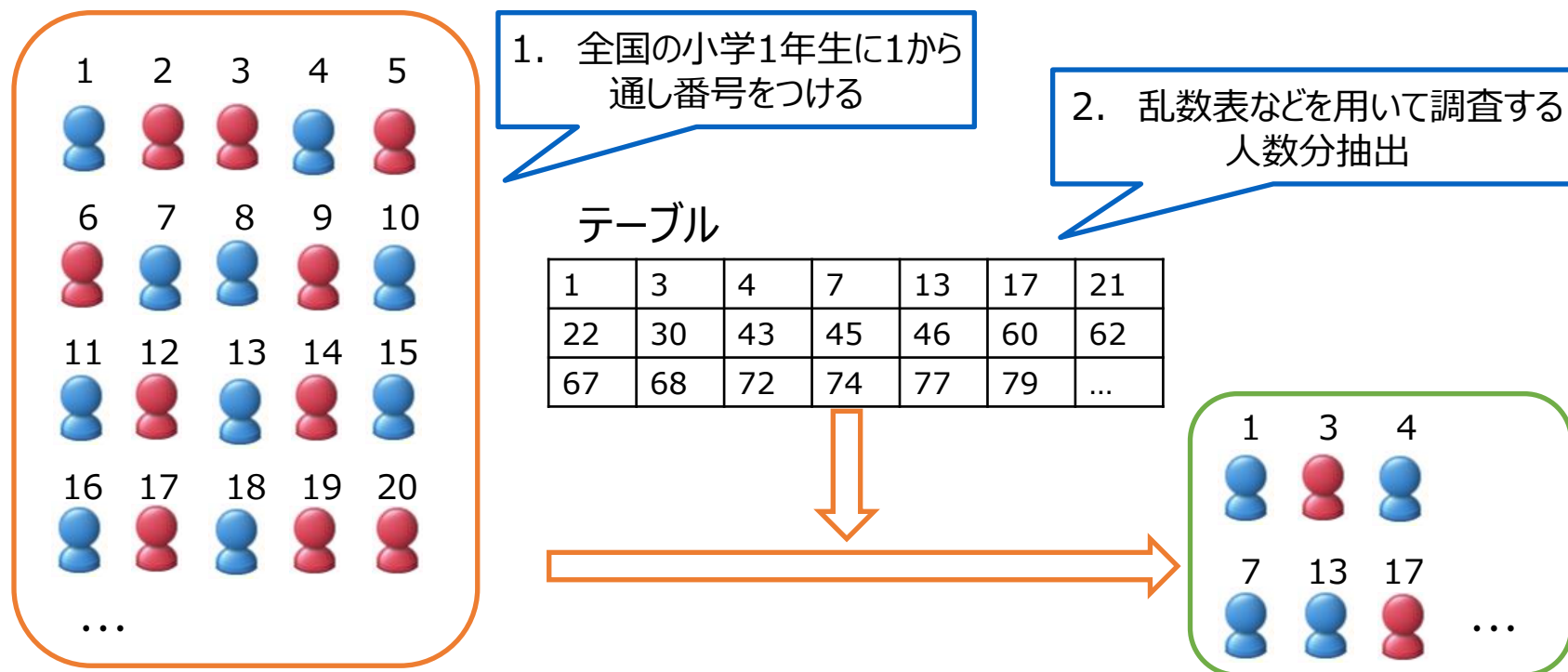
- 母集団から「完全に」・「ランダムに」標本を抽出する方法
- 手順
 - 母集団を準備（これができるか？）
 - 母集団に通し番号を振る
 - 乱数表（サイコロのようなもの）などを用いて標本を抽出する



参考・図 <https://bellcurve.jp/statistics/course/8005.html>

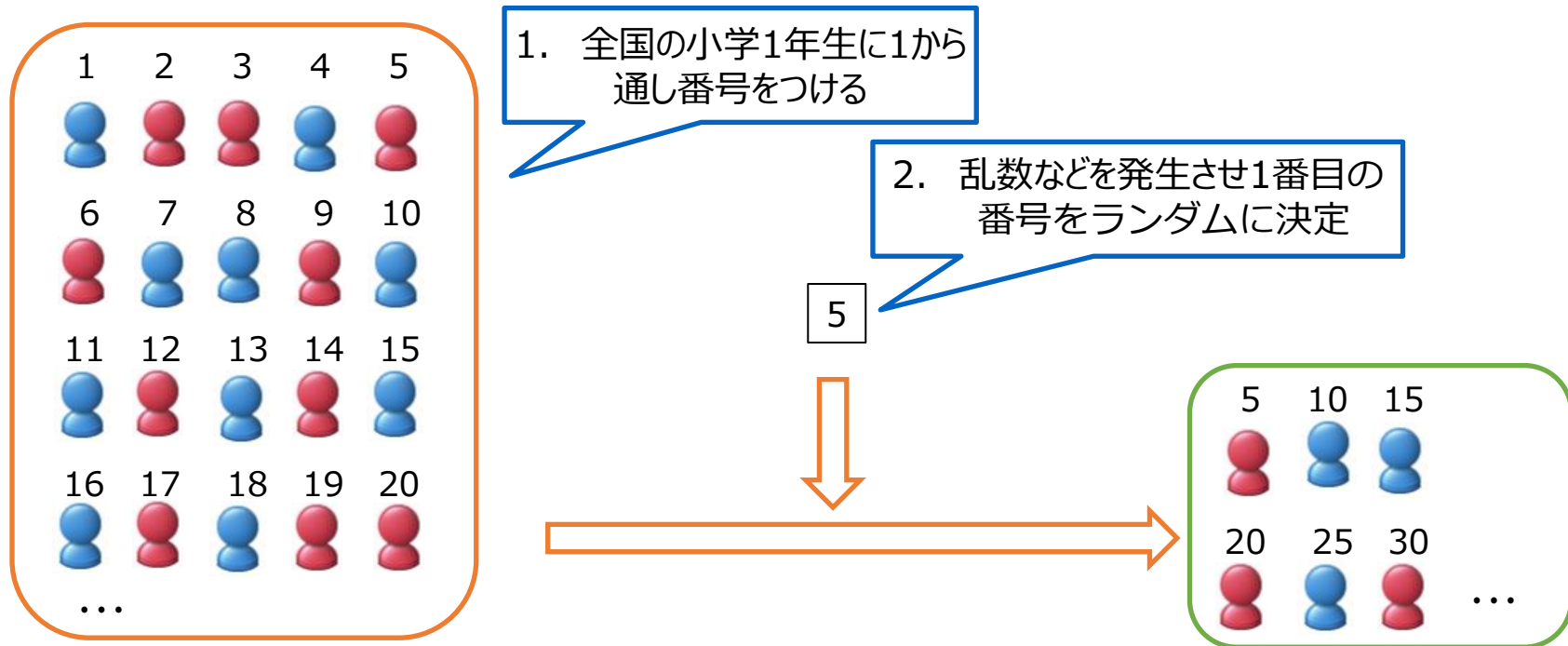
単純無作為抽出の例

●例) 小学生1年の平均身長の調査



系統抽出法

- 母集団に通し番号を振り、1番目の標本を無作為に選んだ後、2番目以降の標本を一定の間隔で抽出する方法



クラスター抽出法 (集落抽出法)

- 母集団をいくつかのグループに分け，その中から無作為抽出で選ばれたグループに含まれるすべての標本を抽出する方法
 - 例) 小学1年生の平均身長の調査

各小学校を1グループとする



小学校を無作為抽出

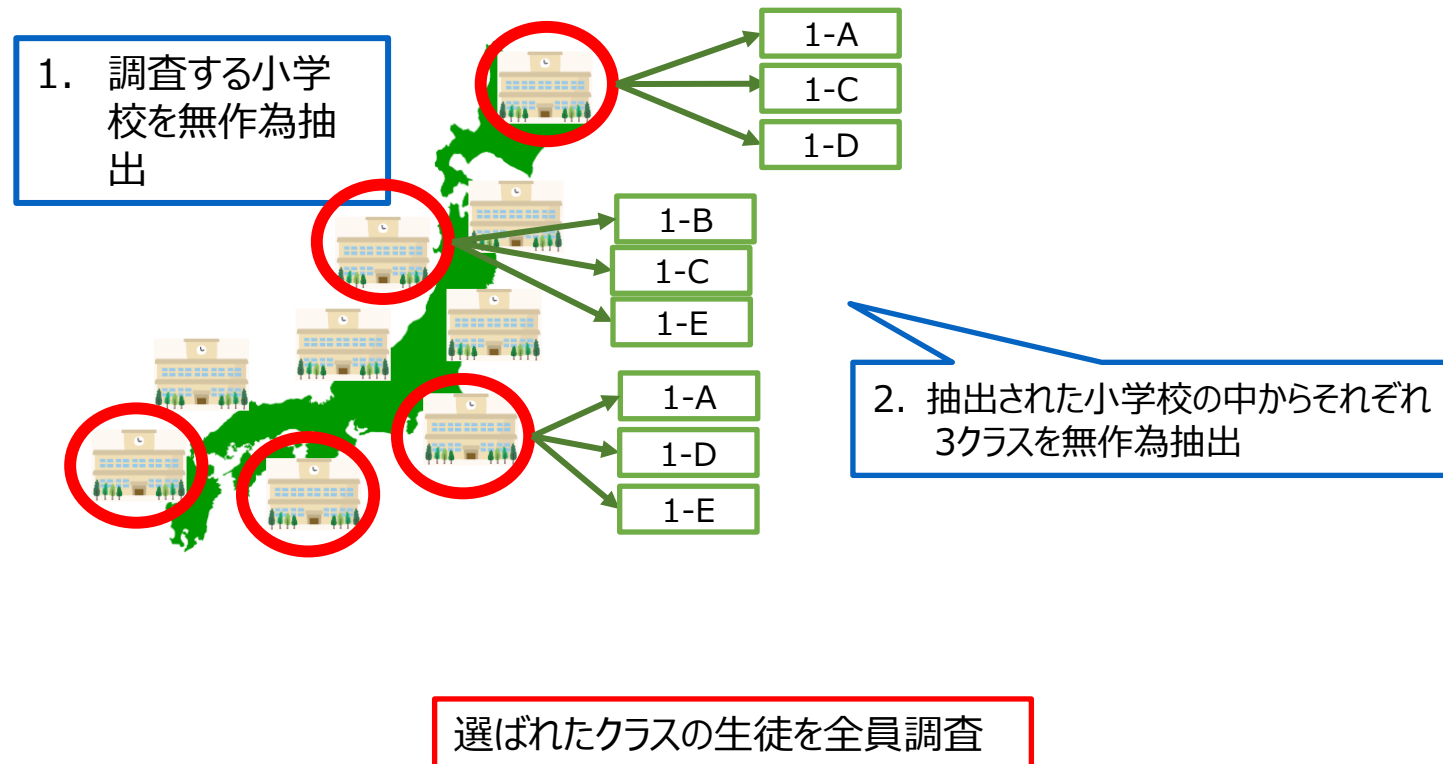


選ばれた小学校の
全生徒を調査



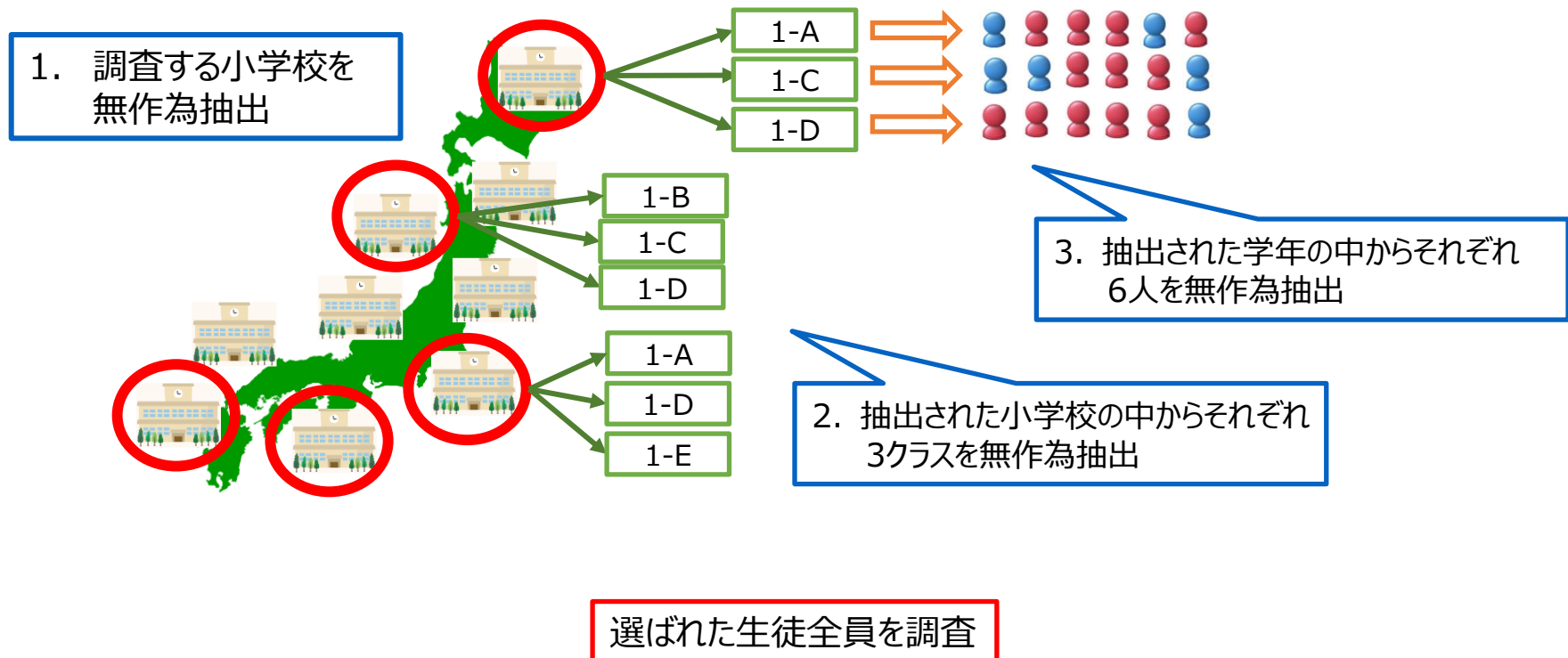
二段抽出法

- 標本の抽出を二段階で行う方法
 - 例) 小学1年生の平均身長の調査



多段抽出法

- 標本の抽出を複数の段階で行う方法
 - 例) 小学1年生の平均身長の調査



層化抽出法

- 母集団が何らかのグループに属しているとき、各グループからその比率に基づいて標本を無作為に抽出する手法

例) 小学1年生の平均身長の調査



地方別小学校数

北海道地方	1061
東北地方	1924
関東地方	4145
中部地方	3489
関西地方	3246
中国地方	1529
四国地方	887
九州・沖縄地方	2927

各地方の学校数の比率と比例する
抽出数を各地方に割り当てる



北海道地方	53
東北地方	96
関東地方	207
中部地方	174
関西地方	162
中国地方	76
四国地方	44
九州・沖縄地方	146

無作為抽出法

- 調査の企画者(ら)の主観的判断を排除して,
母集団からできるだけバイアスが無いように標本抽出
- 母集団から直接的に標本を抽出する方法
 - 単純無作為抽出法
 - 系統抽出法
- サンプルングに基づいて標本を抽出する方法
 - クラスター抽出法 (集落抽出法)
 - 二段抽出法 (多段抽出法)
 - 層化抽出法

層化二段抽出法

● 層化抽出と二段抽出を組み合わせた手法

例) 小学1年生の平均身長の調査

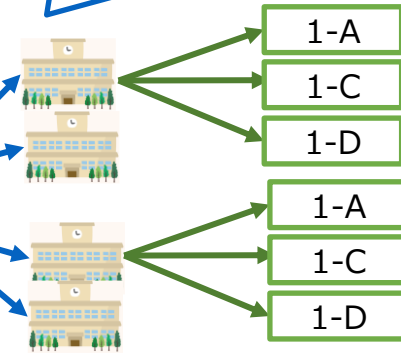
1. 各地方の学校数の比率と比例する
抽出数を各地方に割り当てる

北海道地方	1061
東北地方	1924
関東地方	4145
中部地方	3489
関西地方	3246
中国地方	1529
四国地方	887
九州・沖縄地方	2927



北海道地方	53
東北地方	96
関東地方	207
中部地方	174
関西地方	162
中国地方	76
四国地方	44
九州・沖縄地方	146

2. 各グループから無作為に小学校を抽出



3. 抽出された小学校の中からそれぞれ
3クラスを無作為抽出

選ばれたクラスの生徒を全員調査

各抽出法のメリット・デメリット

	メリット	デメリット
単純無作為抽出法	抽出に偏りがほとんど無い 母集団のサイズだけを把握しておけば実行できる	母集団のサイズが大きい場合、非常に手間と時間がかかる
系統抽出法	単純無作為抽出より手間や時間やコストがかからない	調査対象の並び順に何らかの周期があると標本に偏りが生じる可能性がある
クラスター抽出法	クラスター情報(例: 高校名など)があれば抽出することができるため、時間や手間を節約できる	同じクラスターに属する調査対象は似た性質を持ちやすいため、標本に偏りが生じる可能性がある
二段抽出法 (多段抽出法)	コストを低く抑えられる 抽出効率が高い	サンプルサイズが小さい場合、標本に偏りが生じる可能性がある
層化抽出法	母集団内の情報(例: 年齢別, 性別)を比較することができる 母集団の推測精度が高くなる 各層の分布が大きく異なる場合に使うことができる	母集団の構成情報を事前に把握しておかなければならない

個人情報保護法の動向



個人情報保護法制定 (2003成立, 2005施行) の背景

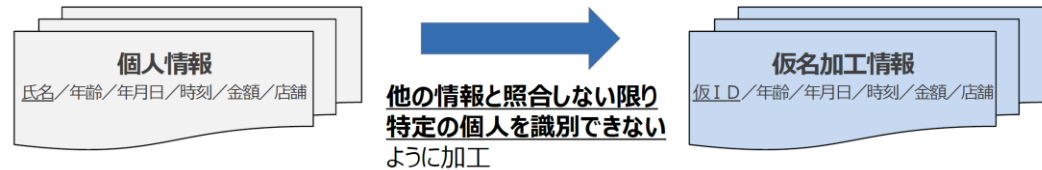
- 情報化社会の進展とプライバシー
 - 私生活をみだりに公開されないというプライバシーの権利が、自己情報コントロール権を含むものへ
- OECD(経済協力開発機構)が、1980年にプライバシー保護と個人データの国際流通についてのガイドライン(通称、OECD8原則)を発表
- 欧州連合の個人情報保護指令
 - 第三国へのデータ移転に、適切な保護提供を求める
- 住基ネットの運用開始(2002年)

個人情報保護法改正 (2015成立, 2017施行) の背景

- 近年の飛躍的な情報通信技術の進展にともない生じた保護すべき個人情報の範囲の精緻化
 - 個人識別符号という考え方を導入
 - 個人情報に該当するか微妙なグレーゾーンの解消を図る
- 個人情報の適切な利活用の手続き
 - ビッグデータの利活用のため匿名加工情報という考え方を導入
 - 一定の条件下で第三者提供が可能に
- 国境を越えた個人情報の流通への対応
 - OECDガイドライン2013年版

個人情報保護法改正 (2020成立, 2022施行) の背景

- 漏洩報告の義務化
- 不適正な方法による利用の禁止
- 「**仮名加工情報**」制度の新設



(参考) 個人情報・仮名加工情報・匿名加工情報の対比 (イメージ)

	個人情報※ 1	仮名加工情報※ 2	匿名加工情報※ 2
適正な加工 (必要な加工のレベル)	—	- 他の情報と照合しない限り特定の個人を識別することができない - 対照表と照合すれば本人が分かる程度まで加工	- 特定の個人を識別することができず、復元することができない - 本人が一切分からない程度まで加工
利用目的の制限等 〔利用目的の特定、制限、 通知・公表等〕	○	○ 〔 - 利用目的の変更は可能 - 本人を識別しない、内部での分析 - 利用であることが条件 〕	× (規制なし)
利用する必要がなくな ったときの消去	○ (努力義務)	○ (努力義務)	× (規制なし)
安全管理措置	○	○	○ (努力義務)
漏えい等報告等	○ (改正法で義務化)	× (対象外)	× (対象外)
第三者提供時の 同意取得	○	— (原則第三者提供禁止)	× (同意不要)
開示・利用停止等 の請求対応	○	× (対象外)	× (対象外)
識別行為の禁止	—	○	○

※1：個人データ、保有個人データに係る規定を含む。 ※2：仮名加工情報データベース等、匿名加工情報データベース等を構成するものに限る。

国際的な動き

- EU一般データ保護規則 (GDPR)

- EU(欧州連合)内における個人情報の保護に関する統一的なルール. いわば個人情報保護法の欧州版
- 原則として個人情報の域外への持ち出しは禁止されるが, 日欧間では可能に(2019年から)

- 忘れられる権利

- ネット上の個人情報や誹謗中傷を削除してもらう権利のこと
- ネットの普及により知られたくない過去の情報が爆発的に拡散し, プライバシ侵害や名誉棄損に苦しむ人が増えてきたことから生まれた
- 過度に重視すると、表現の自由や国民の知る権利を損なうため, バランスが重要

統計法

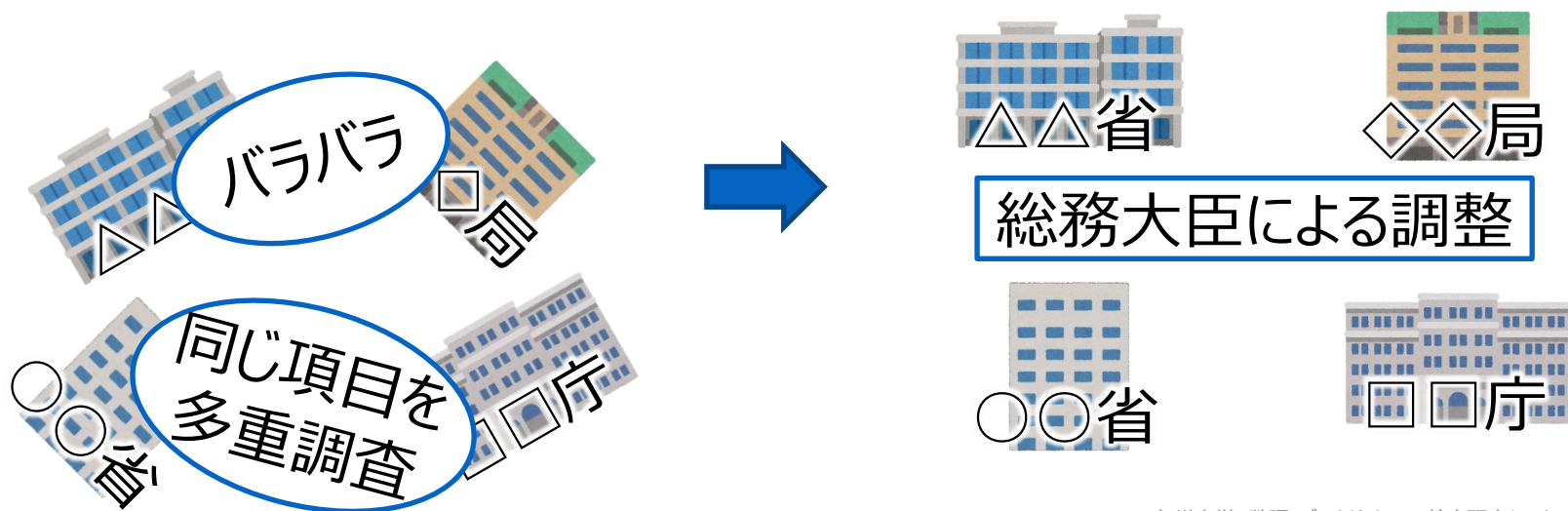
国勢調査など「国全体としての統計」に関する法律

総務省「統計法について」 http://www.soumu.go.jp/toukei_toukatsu/index/

統計局「なるほど統計学園」 <https://www.stat.go.jp/naruhodo/>

公的統計の総合的・計画的整備の推進

- 「公的統計の整備に関する基本的な計画」（基本計画）の策定
 - 有識者会議やパブリックコメントを踏まえた上で、閣議決定
- 【旧】 各府省がその所掌に応じてそれぞれ統計を作成
 - ↓
 - 【新】 政府一体として体系的に整備



統計調査の対象者の秘密保護の強化

- プライバシーの配慮
 - 情報を目的外利用・提供した者や、守秘義務規定違反者への罰則を整備強化
- 「かたり調査」の禁止
 - 国の統計調査をかたって情報を盗み取ろうとする詐欺行為



統計作成への行政記録の活用促進

- 行政機関が業務で収集・蓄積している「行政記録情報」を、統計の作成に活用
 - 調査を受ける国民や企業の負担減



統計データの利用促進

- 【旧】集計表のみ利用可能



- 【新】学術研究や高等教育に協力

- オーダーメイド集計

- 一般からの委託に応じて既存の調査票から新たな集計表を作成・提供

- 匿名データの提供

- 調査票情報の提供

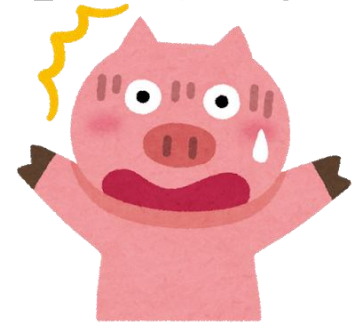
- オンサイト利用：許可を受けた大学・研究機関等から調査票情報にアクセス

規制の集計表枠組みに縛られない
独創的な分析がしやすくなった！



「個人情報保護法」との関係

- 調査票情報に含まれる個人情報は「個人情報保護法」の適用外
 - 基幹統計調査と一般統計調査の両方とも
 - 情報は、集計後に公開（利用・提供）される
- 統計法で独自に、情報保護規定を定めている
 - 匿名データ作成時に、個人情報が漏洩しないように万全の措置をとること
 - 匿名データの提供を受けた者も、匿名データを適正に管理すること



基幹調査統計

- 特に重要な公的統計
 - 国勢調査（5年に一度、日本に居住している全ての人及び世帯が対象）
 - 国民経済計算
 - その他国の行政機関が作成する統計のうち総務大臣が指定するもの
- 対象者の報告義務
 - 報告拒否や虚偽の報告を禁止
 - 50万円以下の罰金
- かたり調査の禁止
 - 未遂も含め2年以下の懲役又は100万円以下の罰金



本物は必ず
「調査員証」
を着用

(公的) 一般統計・民間の統計

- 基幹調査統計以外の調査に対して、報告義務はない
- さらに民間の調査には、統計法の情報保護規定が適用されない
 - 参加・不参加を、自身や保護者が適切に判断する必要あり



オープンデータの目的の詳細

メリットがいっぱい

① 国民参加・官民協働の 推進を通じた諸課題の解決、経済の活性化

自治体による公共データのオープンデータ化は、地域課題の解決、経済の活性化を促進する有効な手段となります。

先進自治体の事例では、オープンデータ化された保育施設に関する情報を事業者が独自に収集した情報と**マッシュアップ**することで、住民が必要な保育施設情報を容易に収集できるようになっています。

地域課題の解決に向けて官民が現状を共有し、課題を具体化し、その解決策・実現策を一緒に考えるうえで、**データの共有**は欠かせません。公共データがオープンデータになれば、住民、民間団体やNPO、民間企業、教育機関との連携を促進できます。

アイデアソン、ハッカソン等

グループ毎にアイデアを出し合い、解決策をまとめたり、そのためのプログラムを開発したりするイベント等の開催



官民協働の例

公共データを活用したアプリ開発

住みよいまちづくり、防災、観光等の地域テーマのためのアプリ開発に向けた、プログラム開発者や民間企業等と連携



出典：福岡まちあるきオープンデータソン
(<http://www.isit.or.jp/wg8/2014/11/10/datason1/>)

出典：データシティ鯖江 (<http://data.city.sabae.lg.jp/>)

② 行政の高度化・効率化

 全国いずれの自治体も**限りある予算や人員で業務を遂行**しています。行政が地域の課題に対し、今後も継続的に取り組んでいくためには、**業務を効率化し、より優先度の高い業務に注力していく**ことが求められます。

一方、仕事の現場では…

 別々の部署で同じようなデータを、それぞれが作成・管理しているために、二度手間かつ内容に食い違いが生じているケースがあつたりします。

たとえば…

- ・公衆トイレのデータ（観光課、環境課、企画課）
 - ・消防水利のデータ（水道課、消防課）
- など

 オープンデータの取り組みに合わせて、各部局が紙ベースで管理していたデータを**電子化・一元管理**したり、Webサイトへの情報掲載手続きを見直したりすることで、業務の効率化につなげることができます。

③ 透明性・信頼性の向上

👉 政策の立案に用いられた公共データが公開されることは、国民の政策への十分な分析・判断に役立ちます。ついては、行政の透明性、信頼性の向上に繋がります。

先行事例では…

自治体が公開する**予算データ・決算データ**を利用し、自治体内での税金の使われ方を**見える化**した「**税金はどこへ行った？**」というWebアプリケーションを通じて、税金が支える公共サービスの受益と負担の関係をわかりやすく理解できるようになっています。

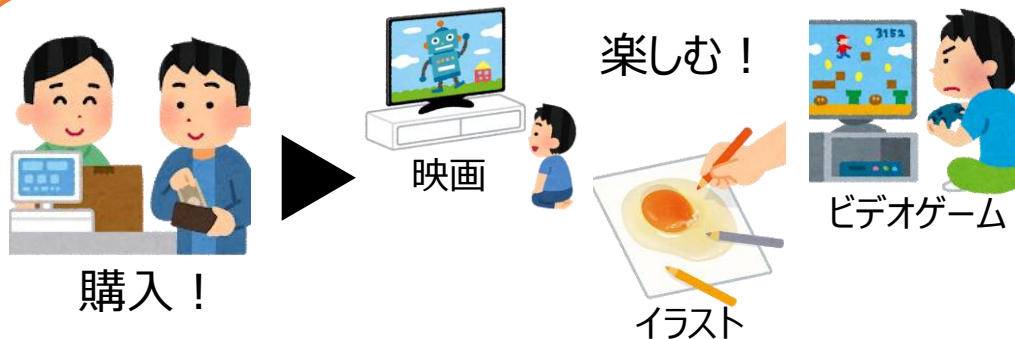
- 更新情報 -	- メディア掲載情報 -
2015/10/02 長野県千曲市が立ち上がりました。	2014/04/03 神戸新聞：「税金の使い道を一目で 尼崎市民らがサイト開設」
2015/10/01 兵庫県加古川市が立ち上がりました。	2013/11/20 ITから考える地方自治 ～「税金はどこへ行った？」をオラの街にも作ろう - Nothing ventured, nothing gained.
2015/06/13 神奈川県川崎市が立ち上がりました。	

→ プロジェクトのこれまでの経歴をすべて見る

出典：税金はどこへ行った？ (<http://spending.jp/#know>)

オープンデータの二次利用と機械判読性

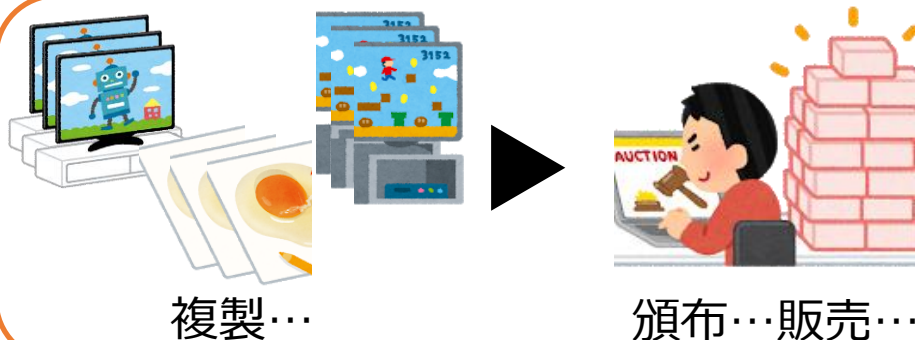
そもそも「二次利用」って何？



自身で買って、
自身で楽しむ！

これは、**一次利用**

※著作物を決められた本来の用途のみに使用する



買ったものを複製して、
配ったり販売したりする！

これは、**二次利用**

※著作物を本来の用途以外に使用する

※著作権者の許可がないと2次利用はやっちゃダメ！法律違反です！
カタチあるものだけではなく、「映画」や「ビデオゲーム」等の「データ」も同様です。

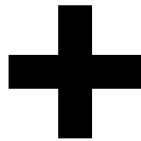
ルールを守り、著作権者の権利を守るのは、とても大事！

でも、なんでもかんでも 「2次利用禁止！」は少しもったいないかも…

たとえば、皆大好きな「ゲーム実況動画」



ゲームプレイ
(1次利用)



動画配信サービス
(YouTubeとか)



ゲーム実況配信
(2次利用)

New!

2次利用によって
「新しいサービス」が
生まれた！
(新たな価値の創造)

※ゲーム会社の許可が大前提！

2次利用の
ガイドラインを
守って配信！
ネタバレは
しないよ！



配信者
(2次利用者)

おもしろ〜い。
続きが気になるし
私も買って
やってみよう！



視聴者
(サービス利用者)

宣伝になって
売上も上がった！？
ネタバレも
されていないし、
ガイドラインを
作っというよかった！



ゲーム会社
(著作権者)

ルールを守った
2次利用で、
皆ハッピーに！

このように…2次利用は新事業（サービス）創出の可能性の宝庫！
でも…だれかが損をしないように、ルール決めがとても大切

「機械判読に適したデータ形式」とは？

利用者が加工してコンピュータプログラムで再利用しやすいデータ形式

使いづらいデータ形式だと…

複雑なデータを処理するのは大変。

データ整理にコストが必要。

緊急事態なのに手入力が必要！



一般的には…PDFなどのバイナリファイル

機械判読に適したデータ形式だと…

アプリケーションを作ろうかな！

ビジネスに使えるかも！

すぐに使える！



一般的には… CSVなどのテキストファイル

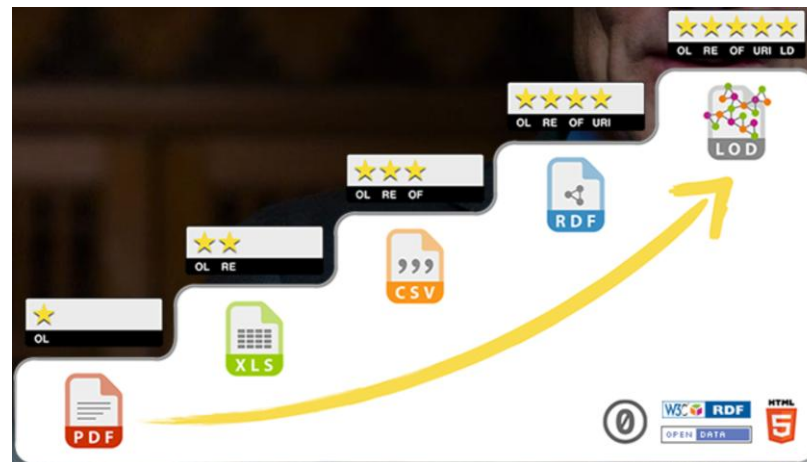
機械判読性のレベル（例）

判読性低



判読性高

LV	種類	形式例
1	OL Open License	.jpg, .pdf
2	RE Reusable	.doc, .xls, .ppt
3	OF Open Format	.csv
4	URI Uniform Resource Identifier	.xml, .json, .rdf
5	LD Linked Data	外部リンクされたLV4



出典：5★オープンデータ (<https://5stardata.info/ja/>)

オープンデータの活用事例

政府CIOポータルオープンデータ100
<https://cio.go.jp/opendata100>

民間の活用事例

No事業名称	概要・テーマ	事業者名	分類
1 アグリノート	農業は、記憶から記録へ	ウォーターセル株式会社	産業創出
2 イーグルバス	センサー乗降データでバス路線の最適化！	イーグルバス株式会社	産業創出
3 カーリル	借りたい一冊、見つかる！	株式会社カーリル	産業創出
4 家計簿・会計アプリ Zaim	もっと、お金に、楽しさを！	株式会社Zaim	防犯・医療・教育等
5 花粉くん	花粉症の日常を、ちょっと新しい体験に。	株式会社 博報堂アイ・スタジオ	防犯・医療・教育等
6 ココゆれ	住宅を建てる前に、地震のリスクがすぐ分かる！	大和ハウス工業株式会社	産業創出
7 さばえぶらり	鯖江の新たな魅力、発見！	ATR Creative	産業創出
8 GEEO（ジージョ）	日本国内の不動産の相場感が分かる！	おたに	産業創出
9 周辺環境スカウター	あなたの住みたい街の戦闘力はいくつ？	30min.（株式会社イー・ド）	産業創出
10 全国避難所データベース	いざという時。すべての人に、避難のための情報を	電通・ゼンリンデータコム	防災・減災
11 ミルモ	介護を支える、新しいカタチ	株式会社ウェルモ	少子高齢、産業創出
12 HalexDream!	気象庁が発表するあらゆるデータを活用	株式会社ハレックス	産業創出
13 セーフティマップ	カーナビデータを活用した事故多発箇所、危険箇所の提供	本田技研工業株式会社	防犯・医療・教育等
14 ワーニング	感染症の流行状況を可視化し、早期対策が可能に！	株式会社オリズン	防犯・医療・教育等
15 全国水利台帳	火災現場で迅速に対応でき、自主防災や地域防災の強化に役立つ！	株式会社placeOn	防災・減災

地方公共団体の活用事例

No	事例名称	概要・テーマ	事業者名	分類
1	会津若松市消火栓マップ	会津若松市消火栓マップ	Code for Aizu	防災・減災
2	かなざわ育なび.net	ワタシだけの、子育て支援ポータルサイト	横浜市 金沢区	少子高齢
3	5374(ゴミナシ).jp	ゴミ捨てスムーズ、いつでもどこでも！	一般財団法人コード・フォー・カナザワ (Code for Kanazawa)	防災・医療・教育等
4	さっぽろ保育園マップ	あせらず、なやまず。パパママまっぷ	Code for Sapporo パパママまっぷチーム	少子高齢
5	鯖江バスモニター	乗りたいあのバスは、今どこ？	鯖江市	産業創出
6	Sabotaつくえなう！	女子高生発、図書館をもっと便利にするアプリ	鯖江市役所JK課	防災・医療・教育等
7	しずみちinfo	静岡市道路通行規制情報をわかりやすく！	静岡市	防災・減災、産業創出
8	税金はどこへ行った？	税金が1日あたりどこにいくら使われているかを知る市民主導のプロジェクト	Open Knowledge Foundation Japan	防災・医療・教育等
9	千葉市民協働レポート (ちばレポ)	あなたの見つけた「困った」が地域の課題を解決する！	千葉市	防災・医療・教育等
10	福島県 帰還支援アプリ	離れていても知りたい、ふるさとのこと。	福島県	産業創出
11	Night Street Advisor	明るい夜道を案内する歩行者ナビゲーションシステム	明石工業高等専門学校 知的情報環境研究室	防災・医療・教育等
12	富岳3776景	共有しよう、「私の富士山」	静岡県・山梨県	産業創出
13	マイ広報紙	市区町村の広報紙をネットやスマホで	一般社団法人オープン・コーポレイツ・ジャパン	防災・医療・教育等
14	室蘭市GIS情報の(一部)オープンデータ化事業	みんなでつくる、街の地図サービス	室蘭市	防災・医療・教育等
15	道路通行実績マップ	災害時道路通行状況を可視化！	特定非営利活動法人 I T S Japan 他、複数の民間事業者	防災・減災