データサイエンス・リテラシー（１）

データの収集

イラスト：©いらすとや

データの収集

Table of Contents

データサイエンスで利用されるデータ

様々なデータとオープンデータ

オープンデータへの大きな流れ

政府データ（オープンガバメント）
オープンソースソフトウェアと集合知データ
科学データ（オープンサイエンス）

データサイエンスにおける倫理

データの種類とその収集方法

みなさんの身近にはどんなデータがありますか？

ロみなさんがアクセスできる（よく見る）「データ」を教えてください.

- 「Webページ」「パソコンのデータ」よりもう１、２段細かなカテゴリで書いてください.

イラスト：©いらすとや

データの種類とその収集方法

収集手段（デバイス）の違いによるデータの種類

ロ 調査(リサーチ)データ

- 研究やマーケティングなどで明確な意図をもって収集
  - ➢ 政府統計、マーケティング、財務状況、アンケート

ロ 観測データ

- 天体観測や気象観測などのように探査機や気象レーダーから収集
  - ➢ アメダス（気象庁）、衛星写真

ロ 実験データ

- 原因の効果を測定するため、他の条件は同じにしたサンプルを作成し収集比較
  - ➢ インターネット環境を使ったオンライン広告の比較実験

ロ 行動ログデータ

- インターネット行動ログやGPSからデータを収集
  - ➢ マーケティング、商圏の分析など

ロ マシンログデータ

- 加工・製造機器やコンピュータから収集
  - ➢ プラントの制御システム
  - ➢ 自動車の制御モニタリング

イラスト：©いらすとや

データの種類とその収集方法

データを獲得する手段

ロ 自分で収集

- 調査、アンケート、スマホログ
- データが集まる仕組みを作る
  - インターネット（アンケートやECサイト）、センサの設置など

ロ 買う

- データ収集を外部に委託
- 他者がすでに収集しているデータ
  - モバイル空間統計@docomo mobaku.jp/price/jpn）

ロ データを持つ会社と提携（会社を買収）

ロ すでにあるデータを活用する

- 自分、所属するグループがすでに持っているデータ
- オープンデータを活用する

一次データ

- 調査者自身が、その調査目的のために固有の方法で採取したデータ.

二次データ

- その調査目的のために採取したものではない、採取済みデータ.
- 収集のためのコストや時間を節約
- 調査者の知りたいデータが揃ってるとは限らない.

オープンデータとは

広く社会に利用してもらうことを目的に公開されたデータ

「一切の著作権、特許などの制限なしで、全ての人が望むように利用・再掲載できるような形で入手できるべきである」

オープンデータの定義（総務省）

- 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
- 機械判読に適したもの
- 無償で利用できるもの

クリエイティブ・コモンズ・ライセンスについて

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」という意思表示をするためのツール

表示（BY）
作品のクレジットを表示すること

改変禁止（ND）
元の作品を改変しないこと

非営利（NC）
営利目的での利用をしないこと

継承（SA）
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

すべての権利の主張

いくつかの権利の主張

すべての権利の放棄

CC BY-NC-ND　CC BY-ND　CC BY-NC-SA　CC BY-NC　CC BY-SA　CC BY　CC0

PUBLIC DOMAIN

PUBLIC DOMAIN

オープンデータの要件

https://opendefinition.org/od/2.1/ja/

1. オープンなライセンス

- 利用、再頒布、改変、分割、編纂の許可 https://creativecommons.jp
  - クリエイティブコモンズライセンスでいえば，CC0, CC-BY, CC-BY-SA

2. オープンにアクセス可能

- インターネットで自由にダウンロード可能，あるいは，実費のみでの配布

3. オープンな形式で利用できること

- １つ以上のフリーソフトで利用、完全に処理が可能な形式
  - エクセル形式よりCSV

データの収集

Table of Contents

データサイエンスで利用されるデータ

様々なデータとオープンデータ

オープンデータへの大きな流れ

- **政府データ（オープンガバメント）**
- オープンソースソフトウェアと集合知データ
- 科学データ（オープンサイエンス）

データサイエンスにおける倫理

オープンガバメント事始め

オバマ政権が2009年1月にオープンガバメントの基本原則を表明

ロ基本三原則

1. **透明性**
   - ➢政府は，国民に対する責任を果たすために，情報をオープンにし，提供しなければならない
2. **国民参加**
   - ➢政府は，知見を広く国民に求め国民の対話を行い，利害関係者グループ外の人々に政策立案過程への参加を促さなければならない
3. **官民連携**
   - ➢組織の枠を超えて政府間および官民連携し、イノベーションを促進しなければならない

オープンガバメント事始め

2009年5月 オープンガバメント施作の第一弾“Data.Gov”を開設

基本三原則
- ●透明性
- ●国民参加
- ●官民連携

政府と国民の
情報の共有が必須

アメリカ政府や自治体がもつ，膨大で貴重なデータをオープンフォーマットやアプリケーション開発に利用できる形式で公開するサイトData.Govを開設

https://www.data.gov/

日本のオープンガバメント

- 2010年5月 内閣高度情報通信ネットワーク社会推進戦略本部が「オープンガバメント」の推進を掲げる
- 2010年6月 同本部がオープンガバメント工程表を発表
  - 但し、10年レベルの詳細不明な工程表
- 2011年3月 東日本大震災に伴う国民の大きな不安
  - 安否に関する情報
  - 放射線情報
  - 電力情報

これらの情報の公開を国民が強く求める

➔政府データのオープン化

日本のオープンガバメント

政府データのオープン化が進んだ例：放射線情報

□政府：「ただちに人体や健康に影響を及ぼす数値ではない」

□国民：「そのうち健康に影響が出るってこと？(不安)」

□各自治体が放射線を測定して実測値を公開

□国民が正しい情報を得る

➔ やったほうがよいこと，しなくてよいことを国民が判断

日本のオープンガバメント

電子行政オープンデータ戦略

ロ2012年7月 内閣高度情報通信ネットワーク社会推進戦略本部が公共データの活用のための戦略を発表

東日本大震災の教訓　https://www.kantei.go.jp/jp/singi/it2/pdf/120704_siryou2.pdf

- ●データが画像で提供されており，機械判読できず人手で再入力する必要がある
- ●行政機関ごとにフォーマットが異なり，情報の収集や整理に多くの時間が必要

世界でも．．．

ロ2013年6月 オープンデータ憲章

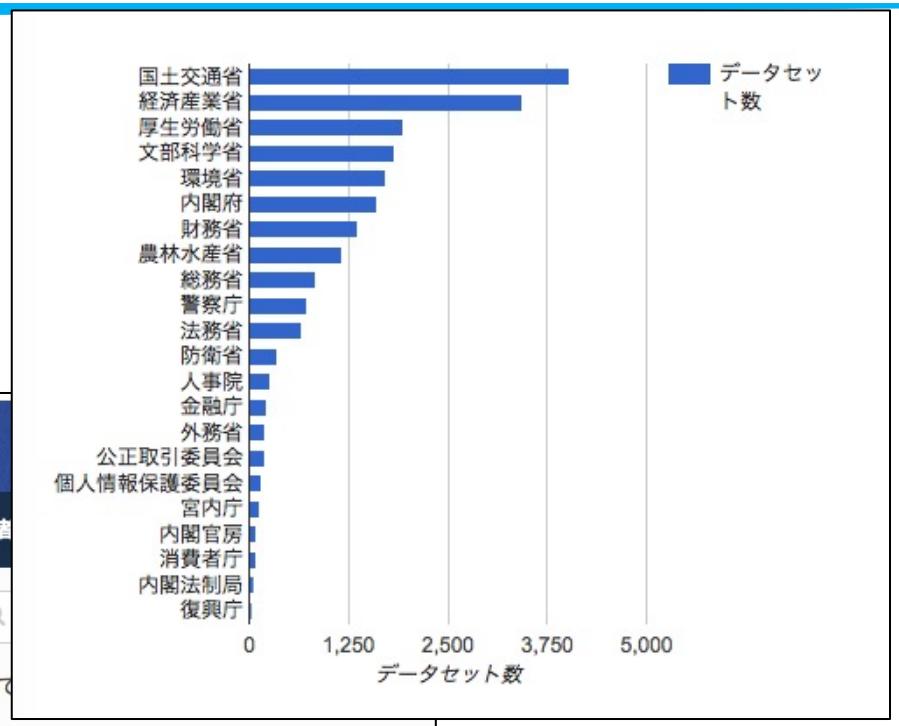
- ●G8(主要国首脳会議)サミットで制定
- ➔ これにより，世界の主要国でのオープンデータ推進が公的に決定

https://www.kantei.go.jp/jp/singi/it2/densi/dai4/sankou8.pdf

オープンな政府データを利用できるサイト

ロデータカタログサイト

- 2014年10月 正式リリース
- Data.Govの日本版



ライセンスはCC-BY

DATA GO.JP データカタログサイト

総務省行政管理局行政情報システム企画課
http://www.data.go.jp/
© Ministry of Internal Affairs and Communications

https://www.e-gov.go.jp/

オープンな政府データを利用できるサイト

ロe-Gov: 電子政府の総合窓口　ライセンスは CC-BY

e-GOV

オープンな政府データを利用できるサイト

https://www.e-stat.go.jp/

□e-Stat

- 日本の統計が閲覧できる政府統計ポータルサイト

オープンな政府データを利用できるサイト

http://opendata-portal.metro.tokyo.jp/www/index.html

ロ東京都オープンデータカタログサイト

ライセンスはCC-BY

東京都、クリエイティブ・コモンズ・ライセンス
表示4.0国際
©Tokyo Metropolitan Government.

データの収集

Table of Contents

データサイエンスで利用されるデータ

フリーソフトウェアの時代

オープンソースソフトウェアへの黎明期

ロGNUプロジェクト

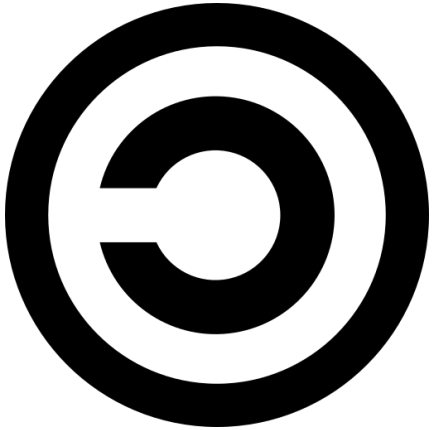
- フリーソフトウェアファウンデーション(FSF)のリチャード・ストールマンが1983年に始めたソフトウェア開発プロジェクト
- 計算機上のすべてのソフトウェア(OS含む)がフリーなソフトウェアだけで済むことを目指した
- GPLライセンスを開発し，公開した

“Blue wildebeest” ©Pcb21
(Licensed under CC-BY-SA 3.0)

GNUのマスコット
“A bold GNU head” ©Aurelio A. Heckert
(Licensed under CC-BY-SA 2.0)

コピーレフト ⇄ コピーライト

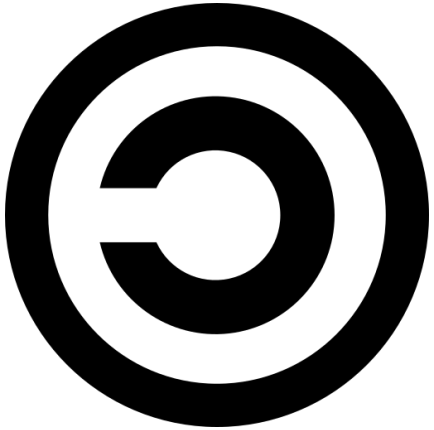
ロGNUプロジェクトにおける フリーソフトウェアについての考え方



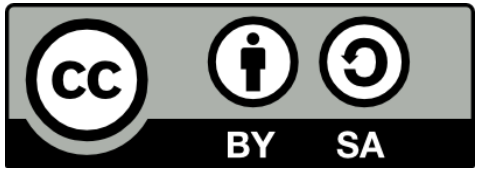
- 著作物の利用，コピー，再配布，翻案を制限しない
- 改変したもの（二次的著作物）の再配布を制限しない
- 二次的著作物の利用，コピー，再配布、翻案を制限してはならない
- コピー，再配布の際には，その後の利用と翻案に制限が無いよう，全ての情報を含める必要がある（ソフトウェアではソースコード含む）
- 翻案が制限されない反面，原著作物の二次的著作物にも同一のコピーレフトのライセンスを適用し，これを明記しなければならない

コピーレフト

ロGNUプロジェクトにおける
フリーソフトウェアについての考え方



●CC-BY-SAはこの考えの流れをくんだもの



●原作者のクレジット（氏名作品タイトルなど）を表示し，改変した場合には元の作品と同じCCライセンス（このライセンス）で公開することを主な条件に，営利目的での二次利用も許可されるCCライセンス.

事実上，企業などが利用し，利益を上げることには向かないライセンス

オープンソース

ロフリーソフトウェアの考え方をゆるめ
ソースの公開と自由な利用に焦点をあてる

- MITライセンス
- BSDライセンス
- Apache2ライセンス　など

➔オープンソースソフトウェアの裾野が広がった

オープンデータの概念と近い概念が現れる

伽藍とバザール

ロフリーソフトウェア時代：

- ソフトウェアの開発は閉鎖的な開発チームにより制限

➔ 安定版でソースコード公開

ロオープンソースソフトウェア：

- ソースコードを不特定多数の利用者・開発者により公開で設計・実装
- 多くの人が実装に関わることで、より迅速に全ての形状のバグを改善
- GitHubのサービス開始(2008年)がこの開発方法(バザール方式)の利用に拍車

いろんな人が関わるとより良いものができる

https://github.com/

GitHub

©GitHub

ソースコードの世界最大のホスティングサービス

バザール型開発およびソースの公開がスムーズにできるようなサービス設計

①フォーク
本家のソースをコピーして自分の手元に

②プルリクエスト
改良や機能追加したら本家にお知らせ

③マージ
本家が採用をしたら取り込まれる

集合知

ロ(不特定)多数の人による個々の判断や知識を収集し蓄積し，最終的に個々の知識だけでは創造できなかったより高い次元の知識や最適な解を導き出すこと

ロオープンソースソフトウェア：

不特定多数が開発に寄与(バザール方式)

➡ 集合知の一種.

- 不特定多数が少しずつ寄与することで，何かを作り上げるという文化は，ウィキペディアやオープンデータの発展につながった.

集合知データの例

ロウィキペディア CC-BY

https://ja.wikipedia.org/

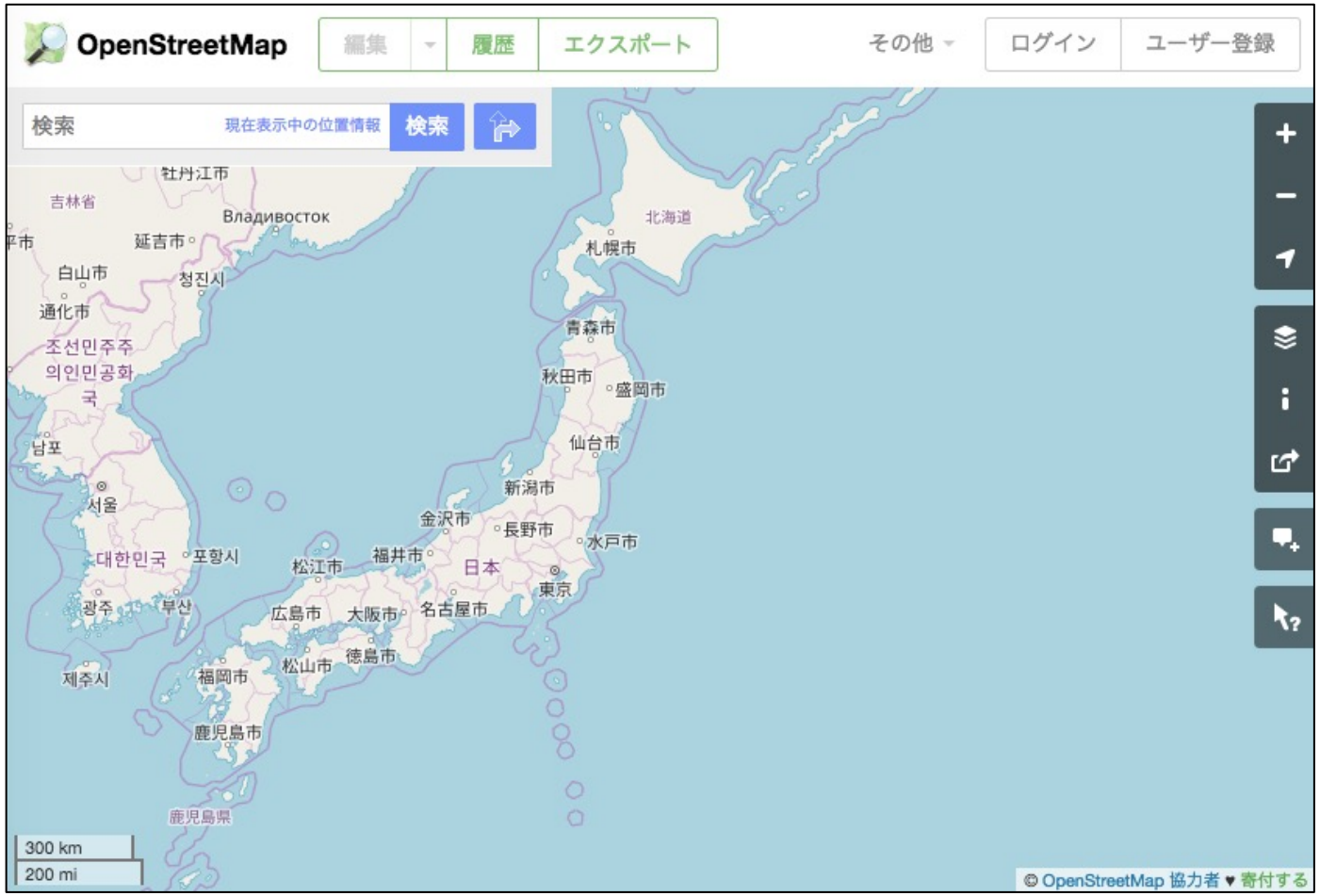
Wikipedia,
A Free Encyclopedia,
Licensed under
CC-BY-SA3.0

集合知データの例

ロオープンストリートマップ

https://www.openstreetmap.org/

ライセンス
ODC-ODbL



OpenStreetMap
(Licensed under CC BY-SA 2.0)

その他の「集合知」を利用したサイト

- **クックパッド：**みんなで作るレシピサイト
  - http://cookpad.com/
- **楽天トラベル：**宿泊予約、クチコミ
  - http://travel.rakuten.co.jp/
- **価格コム：**商品比較、口コミ、レビュー
  - http://kakaku.com/
- **QLife：**医療総合サイト
  - https://www.qlife.jp/
- **食べログ：**グルメサイト
  - https://tabelog.com/
- **CinemaScape：**映画批評空間
  - http://cinema.intercritique.com/
- **@cosme：**化粧品・美容の口コミ
  - http://www.cosme.net/
- **Chakuwiki：**ご当地の噂
  - http://wiki.chakuriki.net/
- **Rakutenみん就：**就職活動情報
  - http://www.nikki.ne.jp/
- **アンサイクロペディア：**誤りと嘘八百でいっぱいの百科事典
  - http://ja.uncyclopedia.info/

・・・・

データの収集

Table of Contents

データサイエンスで利用されるデータ

様々なデータとオープンデータ

オープンデータへの大きな流れ

政府データ（オープンガバメント）

オープンソースソフトウェアと集合知データ

科学データ（オープンサイエンス）

データサイエンスにおける倫理

科学とデータの共有

□ 科学・技術は、先人の成果の上に発見・知見・方法を積み上げた英知

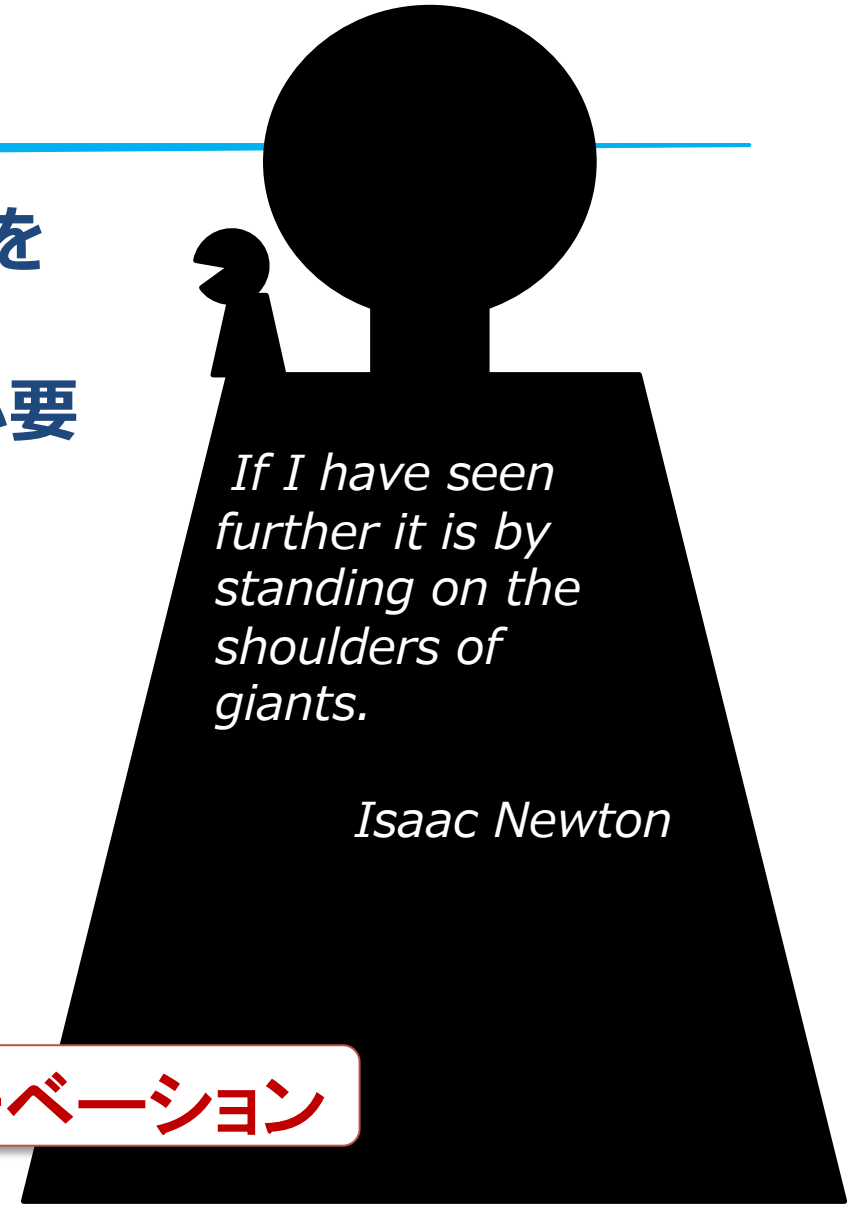
➔ 一方、先人が積み上げた英知にも真偽の検証は必要

論理的には：

- 命題を肯定する証拠を出す場合
  - ➔ すべての場合の証拠が必要
- 命題を否定する証拠を出す場合
  - ➔ 一つの反例が必要

既存の知識に反例が見つかり新しい理論が構築される

データ共有のモチベーション



科学とデータの共有：生命科学の例

□INSDC (International Nucleotide Sequence Database Collaboration: 国際塩基配列協調)

●塩基配列(DNAなど)を使った研究を発表する際には，
日本(DDBJ)，アメリカ(GenBank)，ヨーロッパ(ENA) の
いずれかに**登録済みでないと投稿できない.**

**論文を読んだ他の研究者は，研究に使った塩基配列を
上記データベースから確認できる**

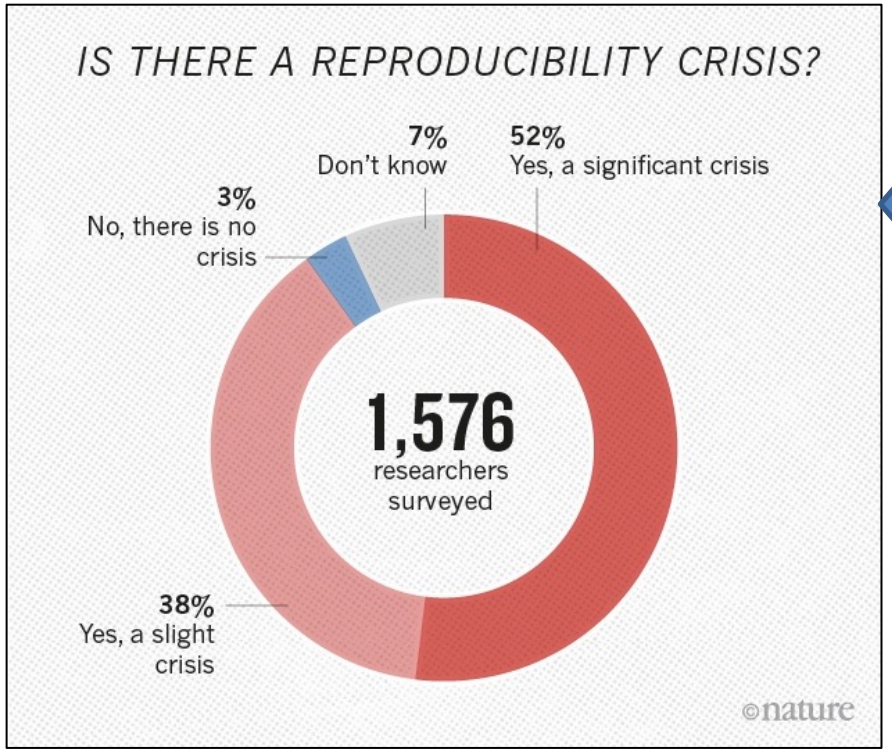
INSDC以外にも広がっている
- wwPDB: タンパク質立体構造
- ArrayExpress/GEO: 遺伝子発現など

イラスト：©いらすとや

再現性に関する調査

ロ科学者1576人を調査.

- 研究者の70%以上がほかの研究者の実験を再現しようと試みて失敗し，半分以上が自分自身の実験を再現することに失敗したことがあると回



再現性の危機はありますか？
の回答の割合　**（90%以上！）**

ロ 解析結果だけではなく，以下も共有

- 実験手法（使った機械，試薬のメーカーなども）
- 生データ
- ソフトウェア（バージョン，パラメータ，ワークフロー）

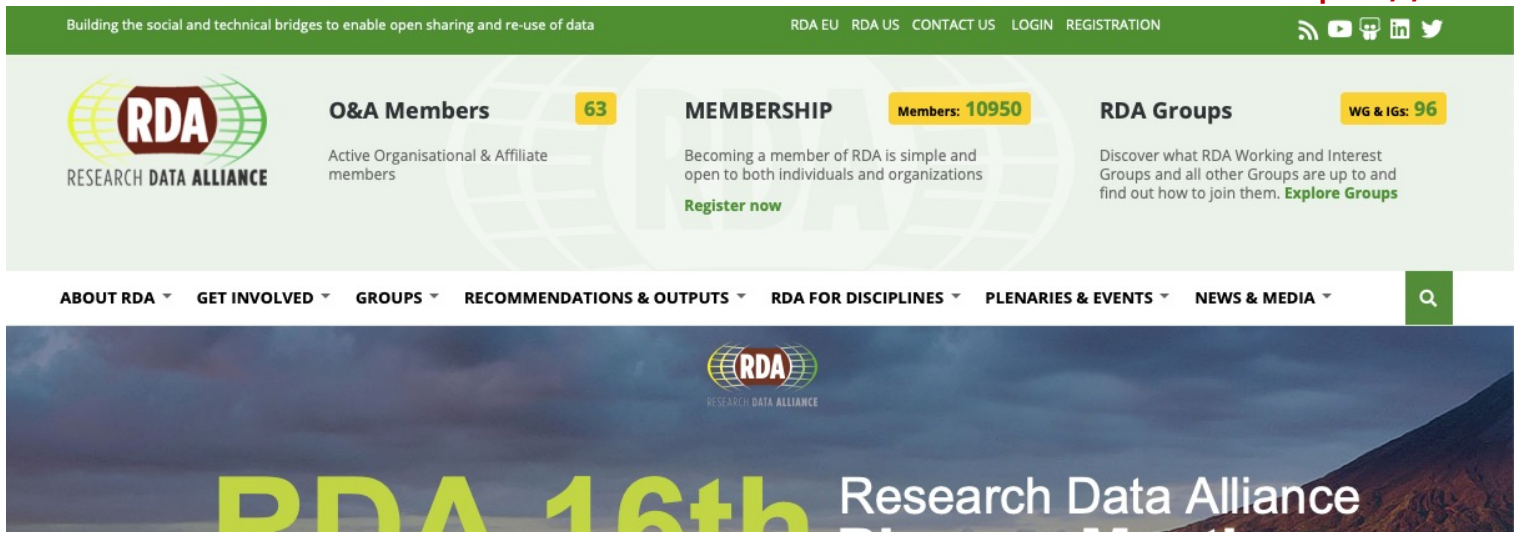
Monya Baker, 1,500 scientists lift the lid on reproducibility Nature, 533 (7604): 452-454, 2016

RDA (Research Data Alliance)

ロ科学データ共有の連携同盟

- 科学研究データの共有を加速し，技術・基盤・応用などを実現していくことが目的.
- 2013年3月発足.　G8が契機.　オープンガバメントと違う形での科学研究オープンデータを目指すための国際組織.
- 実質的な国際標準・国際相互結合体制の形成
  - ➢研究者・技術者・専門家等による合意形成
  - ➢国際的な人材基盤・ノウハウ基盤を他組織と共有して推進。

https://www.rd-alliance.org/



オープンサイエンスの現状

生命科学以外も含めて科学データの公開の機運が高まっている

□生命科学以外の分野でもデータ共有は今後進むと考えられる.

⇔ただし、共有における課題は多い

- ●再現性が担保できるほどの情報はない
- ●個人情報の問題（医療、教育）

□論文による情報共有

- ●オープンアクセス雑誌
- ●arXivなどのプレプリントサーバ

オープンアクセス雑誌とarXiv（アーカイブ）

ロオープンアクセス雑誌

- インターネットを通じて，誰でも無料で論文にアクセス可能である雑誌.
- 論文出版費用は通常著者が全額負担(雑誌によっては，低所得の国の研究者は費用を一部／全額出版社が負担)

ロarXiv（アーカイブ）

- 物理数学計算機科学でのプレプリントサーバ
- 通常の学術雑誌では，数週間～半年／一年の査読期間があるが，このサーバは査読なしで，すぐに掲載
- 研究者同士の評判で論文の価値が判断される(評判の高い論文は，学術雑誌からスカウトされることもある)

論文成果へのアクセスは大幅に向上

arXiv + GitHub によるAI研究の進展

□情報科学以外

●成果→学術論文誌へ提出→査読→出版→再現実験→次の段階へ

□通常の情報科学

●成果→国際会議論文(とGitHub)→短い査読→出版
→(GitHubコードで)再現実験→次の段階へ

□AI研究

●成果→arXiv論文, GitHub
→GitHubコードで再現実験→次の段階へ

回転が非常に早い！

データの収集

Table of Contents

データサイエンスで利用されるデータ

ELSI（エルシー）とは？

新規の科学技術を研究開発し社会実装する際に検討すべき、倫理的・法的・社会的課題

Ethical (倫理的)

Legal (法的)

Social Issues(社会的課題)

代理母はOK?（生殖補助医療）

ドローンはどこでも飛ばして良い？

Twitterでは何を言っても良い？

自動運転はどこまでAIに任せて良い？

- 1988年にJames Watson（生命科学分野）が提唱
  - ヒトゲノム情報が個人と社会に便益をもたらすように利用される仕組みを検討
- ナノテクノロジー、情報技術、原子力技術、コンピュータサイエンス、人工知能（AI）技術などあらゆる科学分野でも研究・検討が広がっている

科学技術が社会に与えうる影響についてどこまで責任を持つべきか？

「倫理的」と「合法的」

倫理的なデータ分析には法的な枠組みを超えた対応が必要

口収集されたデータは様々なバイアス（偏見、歪み）を含む

➔合法的な取り扱い、合理的な分析であっても偏見や差別を助長する結果が得られることがある。

- モデルの法的影響や倫理的影響を慎重に検証
- 人権の尊重や促進に対する利用者の責任と合致しているかを評価

データ取り扱いの健全性（捏造、改竄、剽窃・盗用）

不公正・不適切な研究行為・発表

ロデータ捏造

<実例>

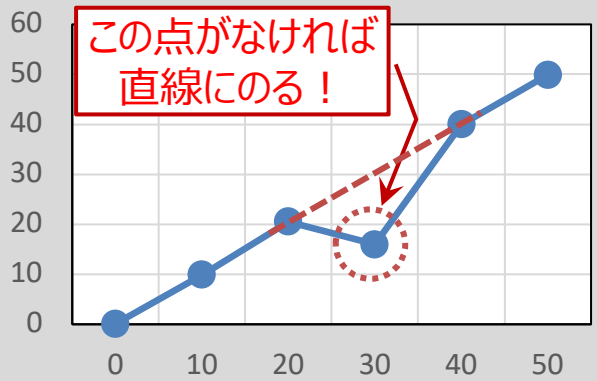
- 有機物高温超伝導
- 転移温度を大幅向上
  - 2000~2001年
- 半導体デバイスの歴史を変える大発明

- 2年に渡る世界中の研究者による再現は失敗
- 異なる論文のデータが使い回されており、捏造されたデータであった。

ロデータ改竄

<文部科学省>

研究資料・機器・過程を変更する操作を行い、データ、研究活動によって得られた結果等を真正でないものに加工すること

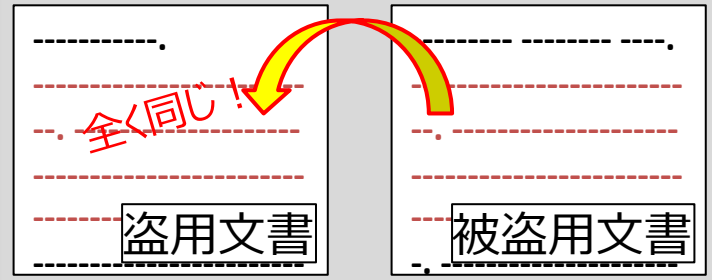


ロデータ剽窃・盗用

- アイデアの盗用
- 文章の剽窃
- 図表を断りなく借用

学問上の犯罪

➔出典明記し正しく引用



CopyContentDetector

剽窃チェッカー

データ取り扱いの健全性（バイアス）

注意が必要なバイアス

体系的な過誤と社会的な偏見

ロ測定バイアス

- 人がデータに間違ったラベルをつけた
- 機械・システムが誤動作した

} データセットにある特定の要因や特性、グループのサンプルが過大・過小になる可能性

ロデータセットバイアス

- データの特性に偏りがある（データが少ないグループは特に注意）

イラスト：©いらすとや

データ取り扱いの健全性（バイアス）

様々なバイアス

ロ 関連付けバイアス

- ステレオタイプに基づくラベル付けによるバイアス
  - 例：青、緑色系の洋服→男の子　赤、ピンク系の洋服→女の子

ロ 確証バイアス

- 仮説を検証する時にそれを支持する情報ばかりを集め、反証する情報を無視または集めようとしないために生じるバイアス

ロ 自動化バイアス

- 自動化されたシステムの結果を信じやすく、矛盾する他の結果は無視しがち

ロ 社会的バイアス

- 社会的な偏見などによって生じた格差によって、生じるバイアス
  - 例：住んでいる地域を分析データに含めることで、意図せず分析結果が人種の違いを反映

データ取り扱いの健全性（バイアスへの対応）

どのようにしてバイアスが入り込むのか？

- **思い込み**（データサイエンティストが持つ偏見・ステレオタイプ）
  - データ収集方法や結果の提供方法などについて思い込みはよくあり、一人で予測・対応は不可能
  - → 研究や設計プロセスの当初から多様な関係者や参加者と取り組む
- **トレーニングデータ**（データセットにバイアス）
  - データの収集対象や過去の限られた情報に縛られる
  - → データの収集範囲等をよく確認
- **モデル**（学習に用いる要因によってバイアスが発生）
  - 例えば、人名から性別、人種、国籍が暗示され分析や予測にバイアスがかかる.
  - → バイアスの影響が懸念されるデータは消去・置換え
- **人為操作**（トレーニングデータ編集によるバイアス除去・付加）
  - トレーニングデータの編集によってバイアスを除去できる場合もあるが付加することもある
  - → 分析結果のフィードバックを随時取り込みモデルを更新する

バイアスが入ったモデルはバイアスをさらに増幅する可能性があるので注意！

データ取り扱いの健全性（バイアスと社会課題）

バイアスはデータだけでの問題ではなく、深く社会に根差している

□「モザイク画像の解像度を高精細にする研究

➔正確にはAIが元画像を類推して「新たな画像を生み出す」技術.

- オバマ元大統領の低解像度写真を全く違う白人の画像に変換.

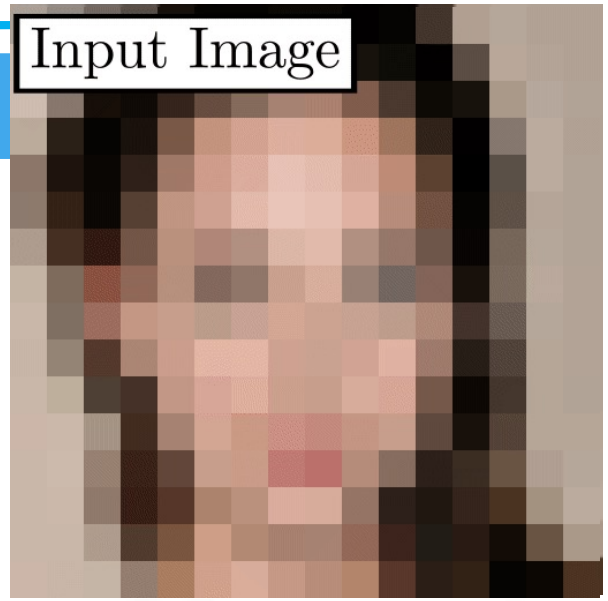
□ Y.LeCun: Facebook人工知能部門チーフ研究者

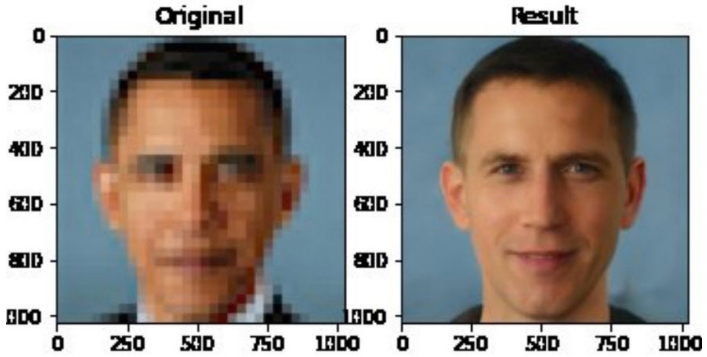
- 白人の写真を多く含むFlickrFaceHQで学習しているためデータにバイアスがかかり白人に見えた.
- データのバイアスの問題であり、機械学習の科学者ではなく、それを利用するエンジニアが注意すべき.

FlickrFaceHQ
白人:　72.6%
アジア人:13.8%
黒人:　10.1%
インド人:　3.4%

□ T. Gebru:「Black in AI」グループの創設者

- 公正さはデータセットだけの問題ではない.
- データのバイアスは、社会の差別構造が反映される.
- 社会に人種や性別への差別があれば、その差別が統計データに現れる.
- さらに、AIの設計、アルゴリズム、実装などの各段階でバイアスが生まれ、強化される可能性がある！





C.Rudin & S.Menon, “PULSE”@デューク大学：
https://github.com/adamian98/pulse

個人情報とプライバシー

イラスト：©いらすとや

□個人情報

- ●名前や生年月日、住所など
  その人が誰なのか特定できる情報
  - ➢個人ID
    - – 住所と氏名（「住所の県名のみ」や「性別」などは単独では個人情報ではない）
    - – マイナンバー（個人ごとに異なる番号）
  - ➢個人の情報・データ
    - – 健康情報
    - – 成績・評価データ
    - – 購買履歴、Webの閲覧履歴なども。。。

居場所、行き先➔最適経路

購買情報 ➔ 欲しい製品の安売り情報

メールアドレス登録 ➔ 不要な広告

➔適切な開示により有益な情報や便益を得られる

➔自分の知らないところで開示されると様々な被害の可能性

権利の保護と情報の有用性の見極めが重要

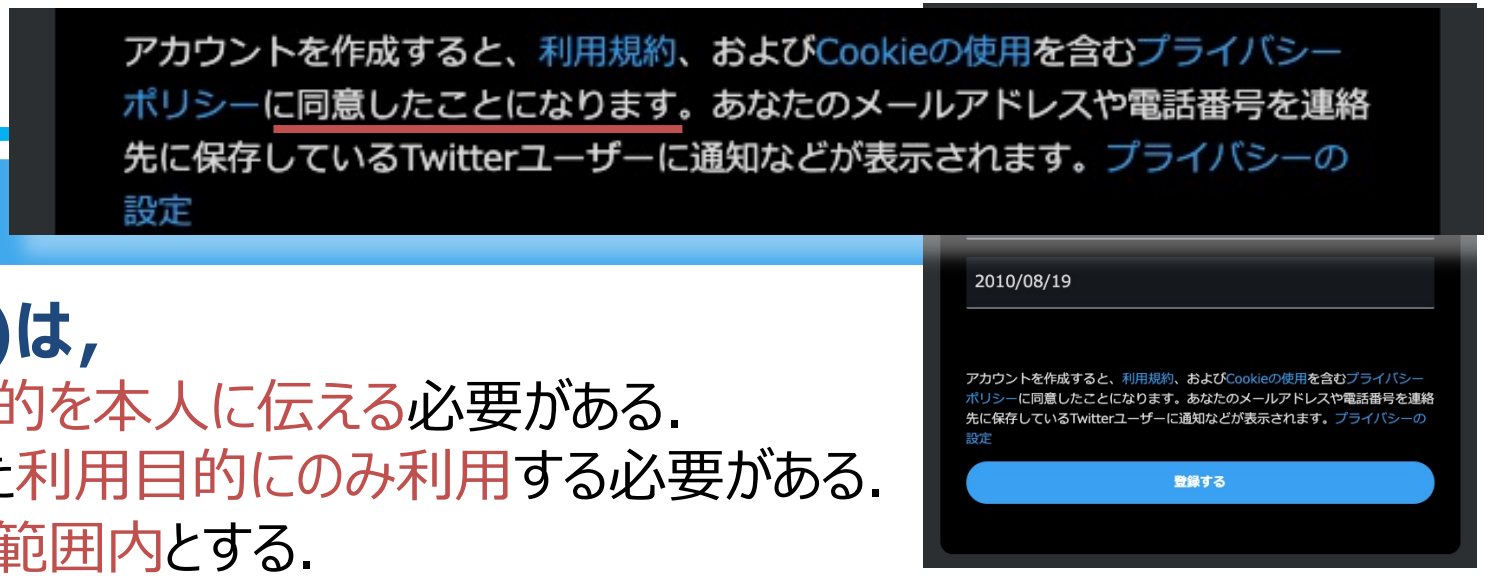
権利の保護　情報の有用性

個人情報・データの保護

情報を保ちながら個人情報を匿名化する様々な手法

- **データ削除**
  利用目的に不要な情報の削除
  – パスポート番号、顔写真、指紋
- **一般化**
  データを上位概念に
  – じゃがいも→野菜、〇〇商事〇〇課→会社員
- **トップ(ボトム)コーディング**
  数値をまとめる
  – 116歳 を 90歳以上、170cmを 150cm以上に
- **ミクロアグリゲーション**
  データをグループ代表値に
  – 年収データを多い方から10人ごとにグループ化し代表値に
- **データ交換**
  ランダムに入換え
  – AさんとBさんの購買履歴を入替え
- **ノイズ付加**
  元の数値にランダムな値を付け加える
  – GPSデータにノイズを加えて自宅や職場特定を困難に
- **擬似データ生成**
  人工データの混ぜ込み
  – データが少なく特異性が高いグループの特定が困難に

個人情報の保護



個人情報保護法の概要

□ 個人情報を取り扱う者(会社等)は,

- 個人情報を取得する際に利用目的を本人に伝える必要がある.
- 個人情報を利用する際には伝えた利用目的にのみ利用する必要がある.
- 第三者への提供は本人の同意の範囲内とする.
- 個人情報は漏洩しないよう厳密に管理する必要がある.

本人に同意を得るのが難しいときや同意をとることに支障がある場合

□ 個人情報第三者提供の例外事項

- 国勢調査など，法令に基づくとき
- 人の生命、身体、財産の保護に必要があるとき（例:事故被害者の家族への連絡）
- 公衆衛生の向上や児童の健全な育成に特に必要があるとき（例:DV, 児童虐待等）
- 国の機関や地方公共団体などの事務遂行に協力するとき（例: 警察の犯罪捜査）
- 個人情報を復元できないよう匿名加工したとき

イラスト：©いらすとや

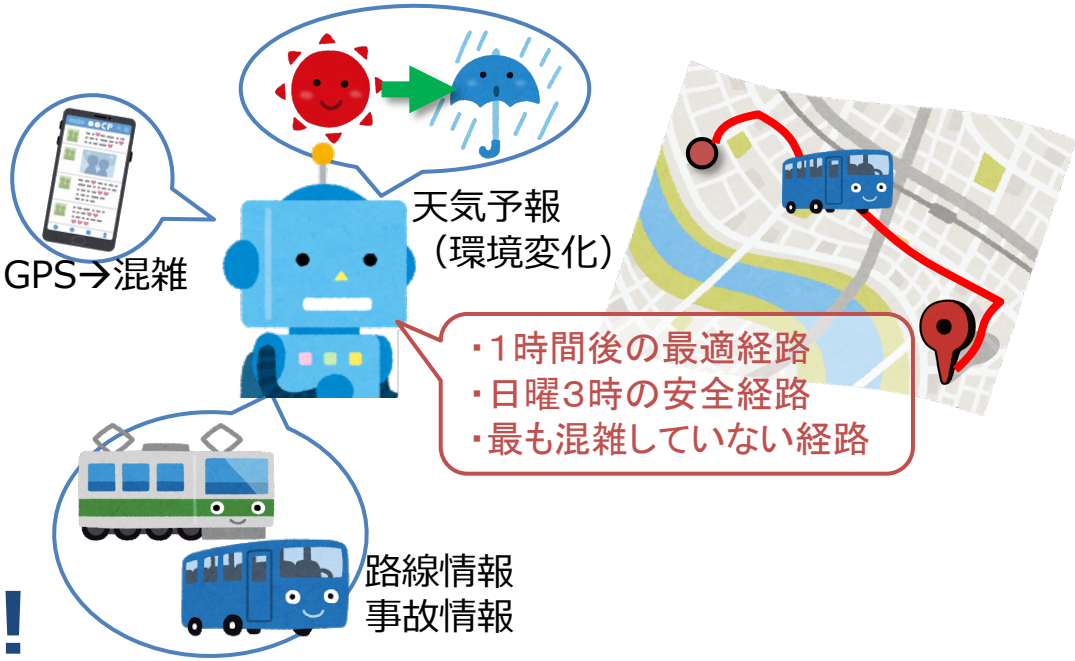
データの保護と活用

匿名化によるメリット

●本人の同意なくデータを利活用可

●事業者間における**データ取引**や
データ連携など、データの利活用を促進

●新事業や新サービスで利便性が向上！



目的にあった匿名化でメリットを得つつリスクを回避

イラスト：©いらすとや

データサイエンスにおける法整備と社会的受容性

GDPR（EU General Data Protection Regulation）

ロ欧州一般データ保護規則（2018年5月～）

- EU内28カ国でバラバラであった個人情報保護関連規則を一元化
- 最先端のデータ保護の考え方を具現化
  - 個人情報・データの所有者は「データ主体」（＝そのデータを発生させた人）
  - 「説明に基づく同意」では目的・期間を必要な範囲に限り，一般人に分かりやすく，いつでも同意を撤回可能
  - AIにより不利な判定が出る場合は，異議申し立てが可能

- 違反した場合：前年度の全世界売上高の4%もしくは2000万ユーロ（1ユーロ125円とすると25億円）のどちらか高い方が制裁金

イラスト：©いらすとや

データサイエンスにおける法整備と社会的受容性

人間中心のAI社会原則（内閣府統合イノベーション戦略推進会議：2019年3月）

（一部抜粋）

- **人間中心の原則**
  - 人が自らどのように利用するかの判断と決定を行う
- **プライバシー確保の原則**
  - 個人の自由、尊厳、平等が侵害されない
  - リスクを高める可能性がある場合には、技術的仕組みや非技術的枠組みを整備
- **セキュリティ確保の原則**
  - 社会は、全体として社会の安全 性及び持続可能性が向上するように務める
  - 、リスク管 理のための取組を進めなければならない
- **公平性・説明責任・透明性の原則**
  - 個人の自由、尊厳、平等が侵害されない

人間中心の AI 社会原則(内閣府ホームページ)：https://www8.cao.go.jp/cstp/aigensoku.pdf