

巨大データが創るデジタルの新潮流： LLM(ChatGPT) / ARW(ロボットによる発見支援) / デジタルツイン

喜連川 優
情報システム研究機構 機構長
東京大学 特別教授
学術会議課題別委員会オーブンサイエンスを推進する
データ基盤とその利活用に関する検討委員会委員長

1

情報システム研究機構とは？



3

2023 G7 Science and Technology Ministers' Meeting



Key points of the Joint Statement

- Key points of the Joint Statement**

 1. G7 Science and Technology Ministers affirmed their commitment to the shared values of democracy, rule of law, and respect for freedom and human rights, as well as the importance of diversity, including gender equality.
 2. In order to contribute to the creation of new knowledge, G7 Ministers need to cooperate in the expansion of **open science** with equitable dissemination of scientific knowledge and publicly funded research outputs including **research data** and scholarly publications following **FAIR principle**.
 3. Further efforts are needed to raise awareness of the risks of unauthorized transfer of knowledge and technology and foreign interference in research and innovation, and to effectively apply mitigation measures as necessary.
 4. G7 Ministers promote international cooperation in outer space, the seas and the Ocean, research infrastructure, international mobility and circulation of human resources to solve global challenges.
https://www2.cao.go.jp/cstp/english/others/2023/g7_2023_en.html

5

变化

- 3月まで 22F
 - 4月から
ROISのオフィスは2F
目の前はドトール



2

COMMUNICATIONS
OF THE
ACM



4

〈科学技術〉

- 40.我々は、グローバルな課題を解決し、次の段階の経済成長を可能にする、イノベーションを推進するための先端技術、研究インフラ及び高技能な人材ネットワークの開発を支援する。このため、国際的な人材の移動及び循環を促進する。G7は、F A I R原則(Findable(見つけられる)、Accessible(アクセスできる)、Interoperable(相互運用ができる)、Reusable(再利用できる))に沿って、科学的知識並びに研究データ及び学術出版物を含む公的資金による研究成果の公平な普及による、**オープン・サイエンス**を推進する。これは、研究者や人々が恩恵を受けるとともに、グローバルな課題に対する知識、イノベーション及び解決策を創造することへの貢献を可能にする。我々はさらに、責任あるグローバルな科学技術協力と、先進的なコンピューティングやバイオテクノロジーなどの新興技術の利用を促進するため、研究とイノベーションにおいて共通の価値観と原則を共有するパートナーとの協力をコミットする。これには、気候変動を踏まえた海洋の更大的理解や、大規模研究インフラの活用が含まれる。我々は、研究セキュリティ及び研究インテグリティ並びにオープン・サイエンスの理念に基づく国際的な共同研究の分野を含め、多国間対話をを通じて、研究及びイノベーションにおける価値観と原則の共通理解の推進並びに促進にコミットする。我々は、予定されているG7パートナーカルアカデミーの立ち上げ並びに研究セキュリティ及び研究インテグリティのベストプラクティスの文書の公表を歓迎する。これらの取組は、安全保障、経済及び科学研究の交際領域において生じる様々な課題への対応に貢献する。

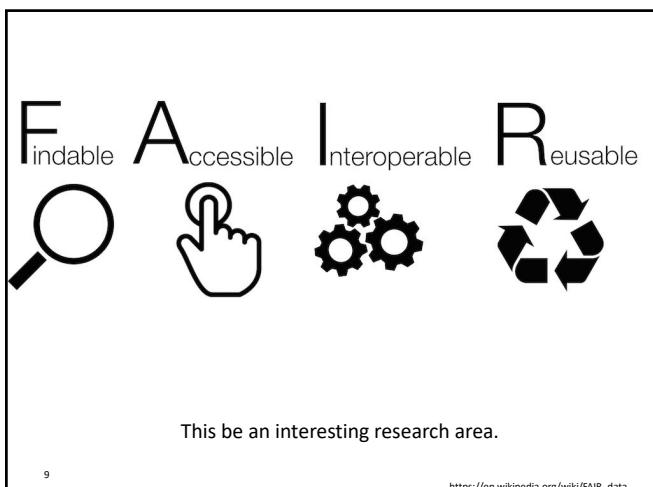
6



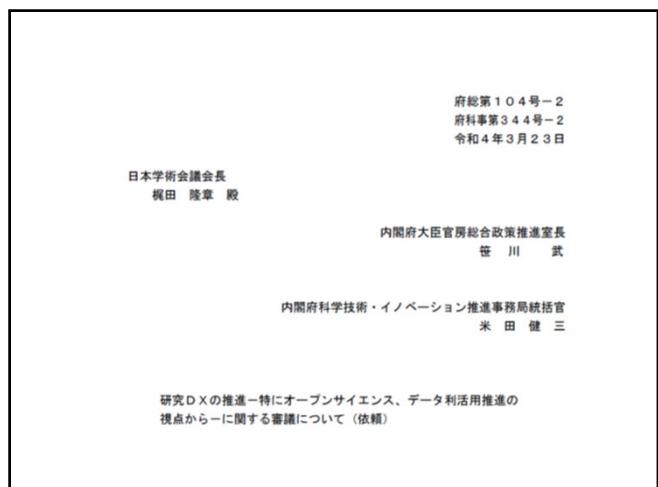
7



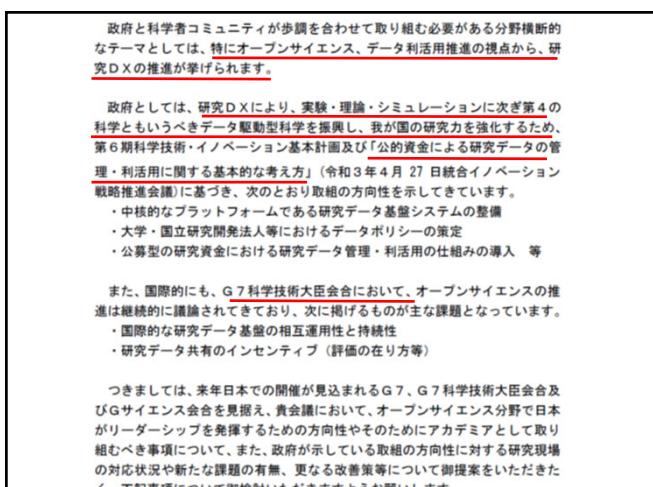
8



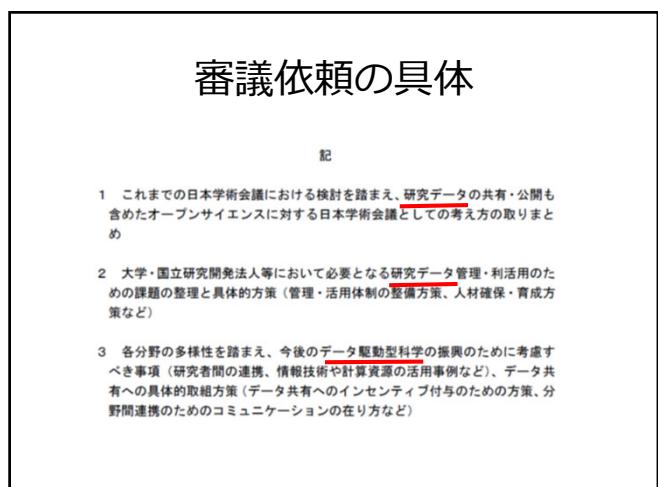
9



10



11



12

日本学術会議オープンサイエンスを推進するデータ基盤とその利活用に関する検討委員会	
委員長	喜連川 優 (連携会員)
	大学共同利用機関主任情報・システム研究機構 国際情報学研究所所長、東京大学特別教授
副委員長	林 和弘 (連携会員・特任)
	文部科学省官能政策研究会データ解析政策研究室長
幹事	三枝 信子 (第三部会員)
	国立研究開発法人立地環境研究所地図システム研究部長
	西田 真也 (第一部会員)
	京都市大学生物学研究科教授
	溝端佐登史 (第一部会員)
	京都市名誉教授、経済研究所特任教授
	小林 武彦 (第二部会員)
	東京大学定年生命科学研究科教授
	小安 重夫 (第二部会員)
	国立研究開発法人理化学研究所理事
	寺崎 浩子 (第二部会員)
	名古屋大学未来社会創造機構特任教授
	仁科 弘道 (第二部会員)
	愛媛大学学長
	藤原 康弘 (第二部会員)
	独立行政法人医薬品医療機器総合機構理事長
	相澤 清晴 (第三部会員)
	東京大学生物学研究科工学系研究所教授
	大橋 弘美 (第三部会員)
	古河電気工業株式会社シニア・フェロー
	筑本 知子 (第三部会員)
	中島大学助教准博士・持続可能なエネルギー研究センター教授
	麦田 公一 (第三部会員)
	明治大学研究・知的財産権特任教授
	秋葉 達也 (連携会員)
	弘前大学研究教授
	木部 梢子 (連携会員)
	大学共同利用機関人間文化研究機構長
	遠澤 実 (連携会員)
	東京農工大学卓越リーダー養成機構特任教授
	高木 利久 (連携会員)
	富士見大学学長
	永井由佳里 (連携会員)
	北里先端科学技術大学院大学院理事・副学長
	大向 一輝 (連携会員・特任)
	東京大学大学院人文社会系研究科相次世代人文学開発センター准教授
	宍戸 喜善 (連携会員・特任)
	東京大学大学院法学政治学研究科教授

13

日本学術会議への審議依頼についての回答

「研究DXの推進特にオープンサイエンス、データ利活用推進の視点からに関する審議について」(2022年12月23日)

内閣府からの審議依頼を受けて、日本学術会議に設置したオープンサイエンスを推進するデータ基盤とその利活用に関する検討委員会、同オープンサイエンス企画分科会及び同オープンサイエンス企画分科会オープンサイエンス・データ利活用推進小委員会が中心となり審議を行った。

【提案1】 研究者が容易に利用可能な研究データプラットフォームの構築

【提案2】 データプロフェッショナルの育成と多面的な研究評価の実現

【提案3】 モニタリング機構に基づくデータ駆動型研究の不断の改善

【提案4】 研究自動化(ARW)に向けた情報技術、計算資源の集約

【提案5】 分野を越えた連携を実現する FAIR 原則の追求

【提案6】 法制度面でのデータガバナンスの構築

※ARW: Automated Research Workflow
※FAIR原則: 「Findable, Accessible, Interoperable, Reusable」の頭文字。
研究データの公開を実現するための理念。

出典: 内閣府科学技術・イノベーション推進事務局が回答から抜粋
https://www.sci.go.jp/a/info/kohyo/pdf/kohyo_25-k335.pdf

14

We are in the age of Data Driven Science NII

Fourth Paradigm 2009/10/1

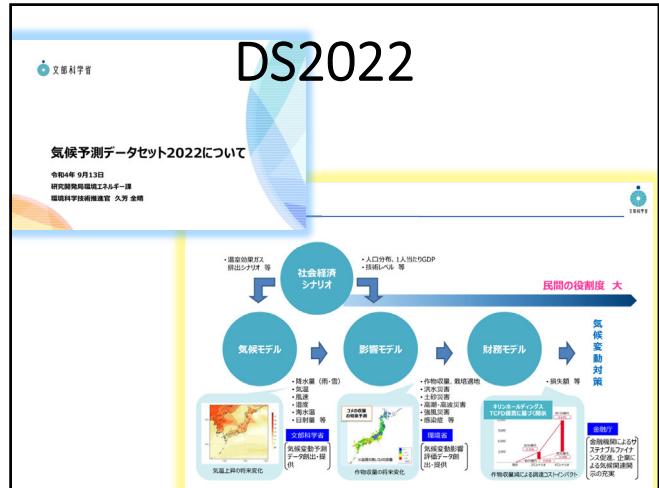
Shift from Computation to Data
Research area with no formula /equations

The Evolution of Science

- Observational Science**
 - Scientist gathers data by direct observation
 - Scientist analyzes data
- Analytical Science**
 - Scientist builds analytical model
 - Makes predictions.
- Computational Science**
 - Simulate analytical model
 - Validate model and makes predictions
- Data Exploration Science**
 - Data captured by instruments
 - Data generated by simulator
 - Processed by software
 - Placed in a database / files
 - Scientist analyzes database / files

Tony Hey
Jim Gray
National Institute of Informatics

15



16

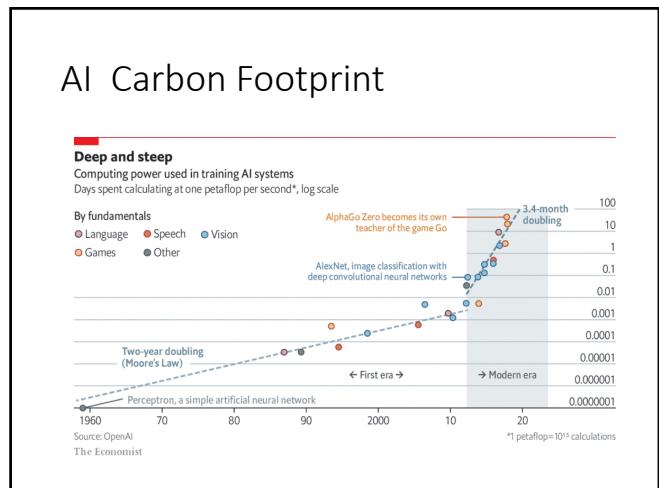
巨大データの時代

気候予測データセットの整備に関する方針

- 我が国の気候変動適応に資する予測情報として
 - ①気候予測データセット
 - ②解説書（予測結果の概要、データ利用ガイド等）
 を整備する。
- これらをデータ統合・解析システム（DIAS）や気象庁ホームページ、気候変動適応情報プラットフォーム（A-PLAT）等に置きユーザーに提供する。

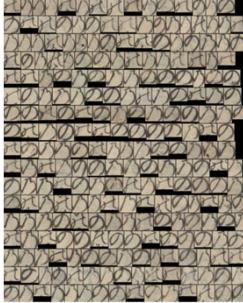
- 約20ペタバイト弱
- 自己のマシンにダウンロード不能
- スパコンで計算するだけではなく（第3）、スパコンで創出されたデータを非スパコンで解析する後工程（第4）が大切になる。
- データプラットフォーム必須の時代へ
(eg. Event attribution) **3000人**

17



18

くずし字 データセット
<http://codh.rois.ac.jp/char-shape/>



Kuzushiji dataset: NIJL and CODH released the dataset with about **4,328** character types and **1,086,326** character images.

Purpose: Develop **machine learning (AI)** models for **kuzushiji OCR**, and use images for human education.

2019/11/22 Digital Humanities in Asian and East Asian Studies 19

19

Data Fuels AI

20

20

審議依頼の具体

記

1 これまでの日本学術会議における検討を踏まえ、研究データの共有・公開も含めたオープンサイエンスに対する日本学術会議としての考え方の取りまとめ

2 大学・国立研究開発法人等において必要となる研究データ管理・利活用のための課題の整理と具体的方策（管理・活用体制の整備方策、人材確保・育成方策など）

3 各分野の多様性を踏まえ、今後のデータ駆動型科学の振興のために考慮すべき事項（研究者間の連携、情報技術や計算資源の活用事例など）、データ共有への具体的取組方策（データ共有へのインセンティブ付与のための方策、野間連携のためのコミュニケーションの在り方など）

21

(2) 大学・国立研究開発法人等において必要となる研究データ管理・利活用のための課題の整理と具体的方策（管理・活用体制の整備方策、人材確保・育成方策など）

- 【提案1】研究者が「**容易に利用可能な**」研究データプラットフォームの構築
- 【提案2】データプロフェッショナルの育成と多面的な研究評価の実現
- 【提案3】モニタリング機構に基づくデータ駆動型研究の不断の改善

22

研究者の責務

6. 研究者の責務

公的資金による研究開発を実施する研究者は、研究成果の最大化や波及効果の誘起、さらには研究成果の利活用を通じて新たな研究成果やノイベーションの創出を促進するために、以下のとおり、適切に研究データの管理を行い、利活用に供する。

6-1. 管理対象データの決定

研究者は、所属機関のデータポリシーや公募型の研究資金における資金記分範囲の基準等に基づき、研究開発の過程で生み出された全ての研究データの中から、管理対象となる研究データの範囲を定める。なお、管理対象となる研究データは研究開発の進捗に応じて、研究者の判断により更新されるものである。

23

研究データの範囲を定める。なお、管理対象となる研究データは研究開発の進捗に応じて、研究者の判断により更新されるものである。

6-2. メタデータの付与

メタデータとは、研究データの第三者による利活用を促進する場合、研究データの管理・利活用に関する情報を記述する場合、日本語の基礎となる書籍情報をとして使用する場合等において重要な情報を指す。

したがって研究者は、管轄ガイドラインにメタデータを付与し、研究データ基盤システム上に記載された情報を元にデータを検索することができる。ただし、研究データ基盤システム上にメタデータが登録され、メタデータが最新版であるかどうかが正確性の最も多くの他のデータプラットフォームを利活用する場合においては、必ず研究データ基盤システムにメタデータを登録しないことを要請する。また、公募型の場合は、公募型の研究資金の管理において必要な場合は、研究者にてメタデータの提出を求めることができる。

6-3. DMP の作成

DMP は研究データの管理計画書であり、研究者による研究データの適切な管理や効率的な研究活動の促進が目的として、研究者や研究プロジェクトをマジメントする者が、資金計画段階のために応じて、原則として研究開始前に作成し、研究の進捗に応じて随時更新するものとする。

6-4. 研究データの保存

研究者は、所属機関のデータポリシー等に沿って、管理対象データを適切に保存することが求められる。特に、管理対象データのうち公開していないデータについては、研究者によつて適正にアクセスされたり、あるいは譲って外部へ漏洩したりすることがないように、十分なセキュリティが確保する必要があります。所属する研究開発を行なう機関等で保管されセキュリティが確保された複数の高いストレージで置かなければなりません。

6-5. オープン・アンド・クローズ戦略に基づく研究データの公開・共有

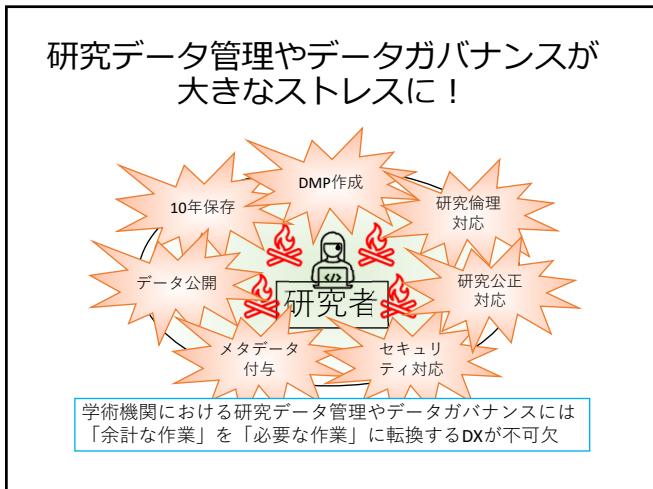
研究者は、管理対象データについて、オープン・アンド・クローズ戦略に従いDMP を策定し、それに基づいて公開範囲を明確化・共有を行う。

6-6. 公募型の研究資金によるプロジェクト等の終了後の取扱い

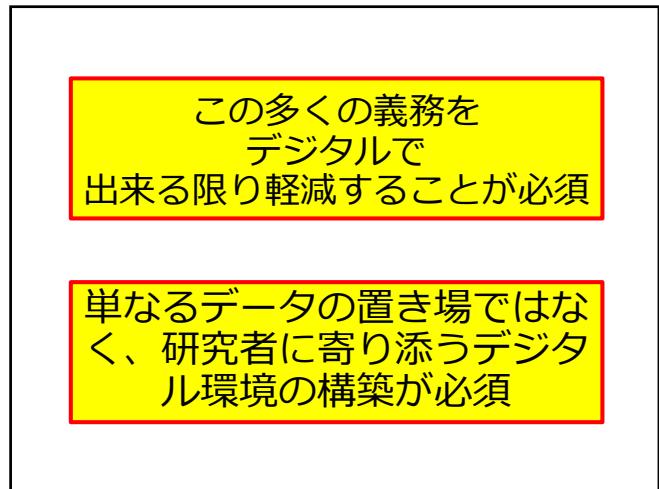
公募型の研究資金等の経理部門からの責任による研究データについては、当該資金の終了時ににおいても、公募型を経由して、研究データが継続的に利活用されることが認められるべきであり、また、不正な使用や不適切な提出を防ぐよう留意すべきである。このため、研究室はあらかじめ、プロジェクトの実施段階から終了後を想定し、期間リポジトリ・分類別

24

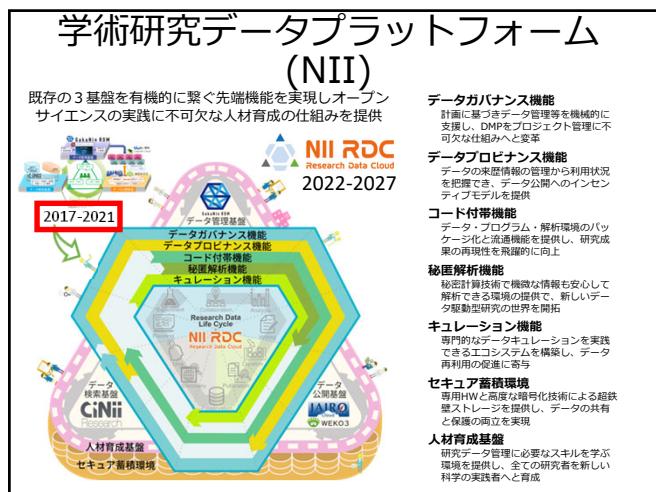
4



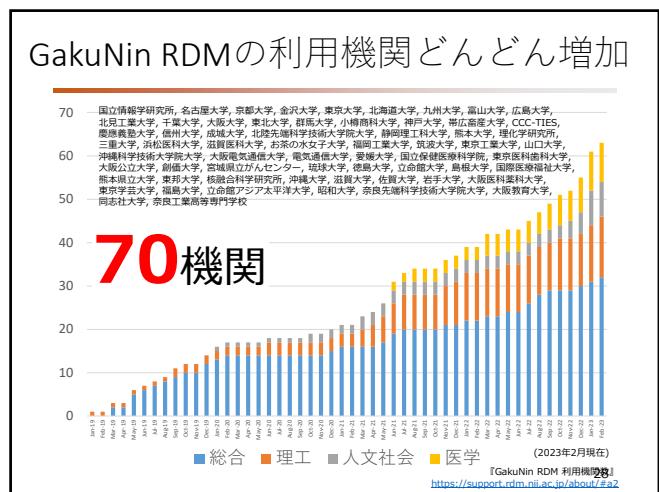
25



26



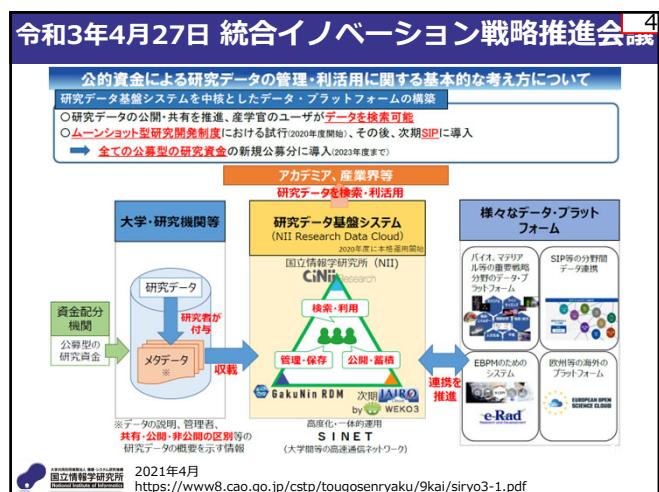
27



28



29



30

DiCE Consortium

■ CINECA（イタリア最大のコンピュータセンター）をコーディネーターとする24のパートナーで構成
■ 欧州各国の研究機関が保有するクラウドインフラを組合わせることで、計50PBを超えるストレージを提供

The DiCE Consortium logo features a blue hexagonal icon with the letters 'DICE' inside, followed by the word 'Consortium'. Below the logo is a grayscale map of Europe with various member institutions highlighted in blue. To the right of the map is a grid of logos for partner organizations, including CINECA, JÜLICH, ASTRON, ICOS, UCL, and others.

31

(2) 大学・国立研究開発法人等において必要となる研究データ管理・利活用のための課題の整理と具体的方策（管理・活用体制の整備方策、人材確保・育成方策など）

- 【提案1】研究者が容易に利用可能な研究データプラットフォームの構築
- 【提案2】データプロフェッショナルの育成と多面的な研究評価の実現
- 【提案3】モニタリング機構に基づくデータ駆動型研究の不断の改善

32

OPENSCIENCE OBSERVATORY

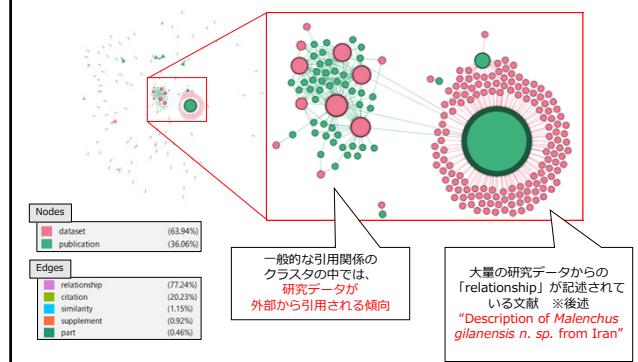
機関ごとに収集した情報を国家・領域ごとに集計。
→研究の活発度やオープンサイエンスへの貢献を可視化

The screenshot shows a map of Europe where each country has a purple circle indicating its publication output. The interface includes a navigation bar with 'HOME', 'OA IN EUROPE', 'METHODLOGY', and a search bar. A sidebar on the right provides an overview of the service's purpose and features. At the bottom, there is a URL 'https://osobservatory.openaire.eu/home' and a page number '33'.

33

OpenAir2

文献や研究データの外部から繋がる関係に重みづけを付加し可視化



34

(2) 大学・国立研究開発法人等において必要となる研究データ管理・利活用のための課題の整理と具体的方策（管理・活用体制の整備方策、人材確保・育成方策など）

• 【提案1】研究者が容易に利用可能な研究データプラットフォームの構築

• 【提案2】データプロフェッショナルの育成と多面的な研究評価の実現

• 【提案3】モニタリング機構に基づくデータ駆動型研究の不断の改善

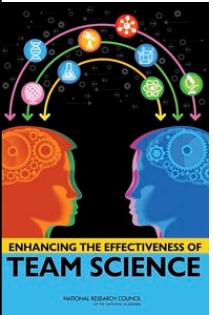
35

データの置き場所に加えて、
データの助っ人が必須

Eg. 電子カルテ

36

Team Science / Data Professionals



• データプロフェッショナルへの新しい評価機構の導入を

• チームサイエンス



National Academy

37

Covid禍でAIツール役立たず？

MIT Technology Review
Featured Topics Newsletters Events Podcasts Sign In Subscribe



Artificial intelligence / Machine learning

Hundreds of AI tools have been built to catch covid. None of them helped.

Sophie Hahn / Getty Images / Getty Images

by Will Douglas Heaven July 23, 2020

38

The Alan Turing Institute



Data science and AI in the age of COVID-19

Reflections on the response of the UK's

39

“ Aspire to a research culture in which data are shared as openly as legal and ethical obligations permit, with central repositories, or 'data lakes', for cleaned and anonymised data ready for analysis.

40

News & Information 順天堂大学

医療・健康

新型コロナウイルスによる肺炎CT画像のAIによる解釈手法開発
～COVID-19肺部CT画像典型症例分類において83.3%の精度を達成～

順天堂大学 Juncture University 名古屋大学 NAGOYA UNIVERSITY

NII 国立情報学研究所

JPS 日本性能評議会

概要

順天堂大学は日本医学放射線学会所属の施設（東京大学、京都大学、大阪大学、九州大学、慶應義塾大学など）の研究代表機関として、東洋大学法、東海薬科大学医療、名古屋大、国立情報学研究所などと共に、新型コロナウイルス（COVID-19）肺部CT画像をAIによって解析する手法を開発いたしました。人工知能技術AIを用いることで、肺部CT画像を入力すると、そのCT画像のCOVID-19肺部典型症例*を判定できるようになります。また、この解析を可能とするために、炎症などによって肺の形状がCT画像上で非常に変形してしまう場合でも、AIが形状に沿った形状を検定できる手法も実現しています。今回開発された手法を用いてCOVID-19肺部CT画像が最も早い結果とそうでない結果の識別において、令和2年1月時点 83.3%程度の典型度測定性能を達成しました。

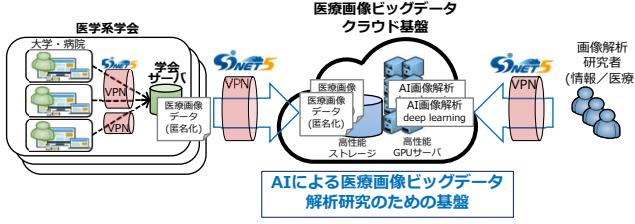
本研究では、日本医学放射線学会と国立情報学研究所を中心とした、AMEDが支援するプロジェクト「医療研究等CT基盤構築・人工知能実装研究事業」「医療ビッグデータ利活用を促進するクラウド基盤・AI画像解析に関する研究」において構築・運用される基盤構造のプラットフォーム上に收集された、我が国におけるCOVID-19病例のデータベースが重要な役割を果たしています。ここでは国立情報学研究所が全国に展開する超高速情報ネットワークSINET5と大阪府・CT画像の収集とAI研究に大変役立っています。今後これらのシステムを活用して、さらなる精度向上が期待されます。

* 医療専門部門が「確実性」の2つで評価するもの。1. COVID-19肺部CT画像の検出率を80%以上のもの。2. 典型度を80%以上とするもの。3. 症状のもの。4. 肺の病変がないものの、これらを区別して、COVID-19肺部CT画像を収集する。

41

国立情報学研究所 医療画像ジャパンワイドビッグデータクラウド基盤 初代AMED理事長 4億枚のパワード

医療画像ビッグデータを解析するための高性能クラウド基盤を開発するとともに、本クラウド基盤を用いてAIによる医療画像ビッグデータ解析技術の研究開発を行い、その利用可能性について検証する。



AIによる医療画像ビッグデータ解析研究のための基盤

42

7

肺野自動セグメンテーション

- 深層学習を用いた3D CT像からのセグメンテーション手法開発
名古屋大学
・肺野の正常領域と炎症領域を識別
- COVID-19症例の肺野状態の3D解析に応用可能

43

放射線学会太鼓判のAI なのに、日本は民転換が 困難！ (アリババが1号)

44

審議依頼の具体

記

- これまでの日本学术会議における検討を踏まえ、研究データの共有・公開も含めたオープンサイエンスに対する日本学术会議としての考え方の取りまとめ
- 大学・国立研究開発法人等において必要となる研究データ管理・利活用のための課題の整理と具体的方策（管理・活用体制の整備方策、人材確保・育成方策など）
- 各分野の多様性を踏まえ、今後のデータ駆動型科学の振興のために考慮すべき事項（研究者間の連携、情報技術や計算資源の活用事例など）、データ共有への具体的取組方策（データ共有へのインセンティブ付与のための方策、分野間連携のためのコミュニケーションの在り方など）

45

(3) 各分野の多様性を踏まえ、今後のデータ駆動型科学の振興のために考慮すべき事項（研究者間の連携、情報技術や計算資源の活用事例など）、データ共有への具体的取組方策（データ共有へのインセンティブ付与のための方策、分野間連携のためのコミュニケーションの在り方など）

- 【提案4】研究自動化(ARW)に向けた情報技術、計算資源の集約**
- 【提案5】分野を越えた連携を実現するFAIR原則の追求**
- 【提案6】法制度面でのデータガバナンスの構築**

46

通常の研究室実験データに加え
多様な大量データ生成源を
有機的に連成

47



48

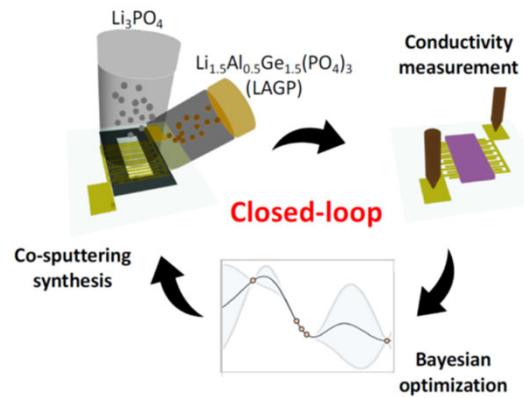
Automated Research Workflow 2022.5

- National Academies



49

Searching new ionic conductors



50



ネットワークを介して自動顕微鏡等の機器やAIサーバーなどと連携し、自律的に判断しながらiPS細胞等の培養実験を遂行。

51

研究の自動化とAIが切り拓く科学と社会 ARW: Automated Research Workflow



52

Data:
Positive & Negative

53



54

ChatGPT/GPT4 左翼、LLaMA右翼

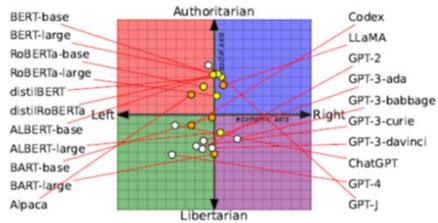


Figure 1: Measuring the political leaning of various pretrained LMs. BERT and its variants are more socially conservative compared to the GPT series. Node color denotes different model families.

55

ChatGPTに乘算をさせてみると

真一 707 × 707 を計算してください。

707 × 707 を計算すると、以下のようにになります:

707 × 707 = 498,049 (不正解)

7 0 7

4 9 4 9

4 9 4 9

4 9 9 8 4 9

真一 本当ですか？

おっしゃる通り、間違いがありました。申し訳ありません。

707 × 707 を正しく計算すると、以下のようにになります:

707 × 707 = 498,049 (不正解)

Shin-ichi Minato



8

56

まとめ



- ChatGPTは「読み書き・そろばん」と言われる能力のうち、「読み書き」だけを徹底的に訓練したAIと言ってよい。
 - 読み書きの能力だけで一定の桁数まで計算できるのはむしろ立派と言える。
- 「そろばん」の能力を習得可能なAIもそのうち（数年以内に？）出てくるのではないか。
 - 計算手順の理解（能力の習得）には、組合せ爆発の問題があり、「教科書」の手順を踏む積み上げ型の学習が必要と思われる。
 - そのためには、「部品」に正しく名前を付ける作法を遵守できるAIが必要と思われる。

57

Parametric vs Non-parametric



Menu

Improving mathematical reasoning with process supervision

58

自動運転解禁 サンフランシスコ 8.10/2023

CALIFORNIA Public Utilities Commission

Contact Us | Translate

Search example: How can I reduce my bill?

Home > News and Updates > All News and Updates > CPUC Approves Permits for Cruise and Waymo To Charge Fares for Passenger Service in San Francisco

CPUC Approves Permits for Cruise and Waymo To Charge Fares for Passenger Service in San Francisco

August 10, 2023 - The California Public Utilities Commission (CPUC) today approved Resolutions granting additional operating authority for Cruise LLC and Waymo LLC to conduct commercial passenger service using driverless vehicles in San Francisco. The approval includes the ability for both companies to charge fares for rides at any time of day.

RELATED PROGRAMS
[Transportation](#)

59

Q4: HOW MANY REPORTED DRIVERLESS AV INCIDENTS IMPEDED SF FIRST RESPONDERS?

Cruise & Waymo do not disclose counts of unexpected stops or other incidents that impeded first responders

SFFD has logged
~50 written reports
of interference in
2023

- **Stops:** Obstructing station ingress/egress
- **Stops:** Obstructing travel to emergency
- **Driving:** Contact or near-miss with personnel or equipment
- **Driving:** Intrusion into SFFD operations
- **Driving:** Unpredictable operations near response zone

City & County of San Francisco: Driverless AV Safety and First Responders

August 7, 2023 | 18

60

10

Q: WHERE AVs ARE BLOCKING SFFD, HOW AND HOW QUICKLY SHOULD SITUATION BE RESOLVED?

AVs should respond to SFFD direction immediately



City & County of San Francisco: Driverless AV Safety and First Responders

August 7, 2023 | 22

- Every second can make the difference between life or death
- A fire's size can double in one minute
- Larger fires lead to > displacement
- "Time is muscle" in response to heart attacks

61

これまでの解説記事

「"やんちゃ"とデータが救う コロナ禍の大学教育」(視点・論点)

2022年01月12日 (水)



国立情報学研究所 所長 喜連川 優



62

2020年3月始め 大規模サイバーシンポジウム約600名 NII

1月初旬 北京出張

2月始め 武漢/北京/上海の研究者との連絡 (春節以降オンライン講義)

2月半ば シンポジウムオンライン化 学会会長決断(学会発表の経験必須!)

3月2~4日 DEIM サイバーシンポジウム開催(多くの見学)

約600人参加 この規模はたまたま日本で初めて
中国でも未着手 (CCF会長とWechat)



63

NII
たった2週間で
本当にやるんですか?
(特段論文にもならないのに)

合田、吉田、横山先生方
+ NIIチーム(柏崎)

64

オンライン化で遅刻が減少

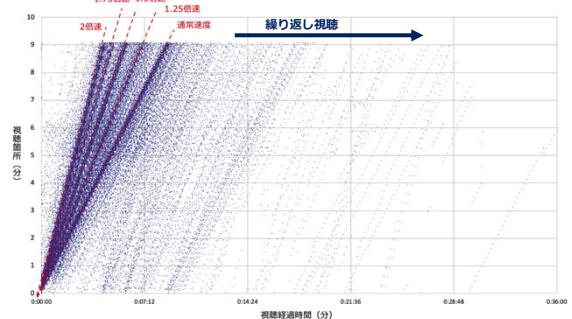


KYUSHU UNIVERSITY

65

大阪大学でも高速視聴が多い2倍速も!

大阪大学初年次必修の一般情報教育科目の第6回講義の動画教材の視聴状況 視聴人数: 2263人



66

11