

2025年4月24日



Research Data Alliance (RDA) 第24回総会参加報告

国大図協資料委員会作業部会委員/
名古屋大学附属図書館 直江千寿子



RDA 第24回総会の概要

- テーマ：「新たなテクノロジーのためのデータ」
- 日時：2025年4月7日～4月11日
- 場所：オンライン
- セッション数：45
- 参加者数：41か国（6大陸）から333人

Africa: 5

Asia: 18

Europe: 205

South America: 22

Oceania: 27

North America: 56



Whova Whova Guides Organizing your own event? About Whova

Research Data Alliance 24th Virtual Plenary Meeting (RDA VP24)

Apr 7 – 11, 2025 | Displaying in your local time: 4:35 PM [Switch to event time](#)

- Home
- Agenda
 - Sessions
 - Speakers
- Attendees
- Community **247**
- Messages
- Photos
- Leaderboard
- Resources
- My Stuff
- Feedback to Whova

Tracks & filters Search session name, location, or speaker name...

Full Agenda My Agenda

< Mon Apr 7 Tue Apr 8 Wed Apr 9 >

4:00 PM

Other - Culture and inclusion

Equity, diversity and inclusiveness in and for RDA activities

Mon Apr 7, 2025 4:00 PM - 5:30 PM **STREAM HAS ENDED** [View details](#) [Add to My Agenda](#)

Location: Whova

Breakout 1

Data Infrastructures and Environ... FAIR, CARE, TRUST - Adoption ... FAIR, CARE, TRUST - Evaluation ...

Which Repository? Community consultation on attributes of repositories & supporting services

Mon Apr 7, 2025 4:00 PM - 5:30 PM **STREAM HAS ENDED** [View details](#) [Add to My Agenda](#)

Location: Whova

Breakout 1

Data Infrastructures and Environ... FAIR, CARE, TRUST - Adoption ... Training, Stewardship, and Data ...

FAIRification processes—achieving FAIR digital assets and services in the life sciences and beyond

Mon Apr 7, 2025 4:00 PM - 5:30 PM **STREAM HAS ENDED** [View details](#) [Add to My Agenda](#)

Location: Whova

Whova（オンライン会議アプリ）を用いて参加

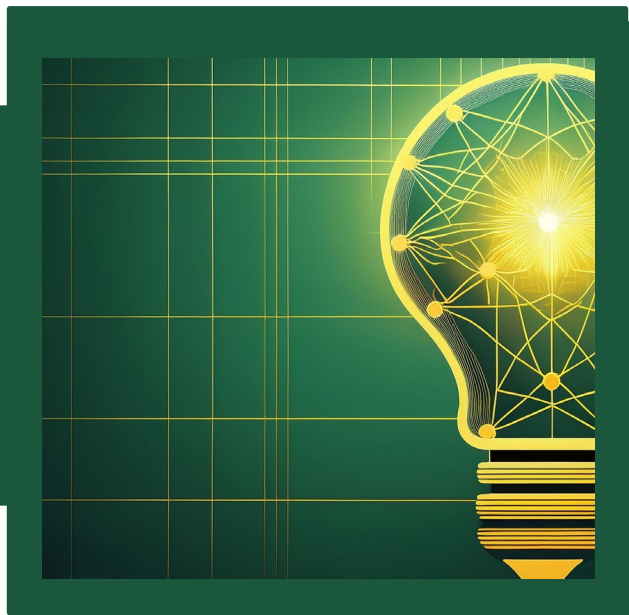


(参考) 各セッションに付与されているタグ

- AI meets data: exploring use cases, applications and innovation
- Data Infrastructures and Environments - Discipline Focused
- Data Infrastructures and Environments - Generalist
- Data Lifecycles – Versioning, Provenance, Citation, and Reward
- FAIR, CARE, TRUST - Adoption and Implementation
- FAIR, CARE, TRUST - Evaluation and Policy
- Other - Culture and inclusion
- Policies
- Research Software
- Scholarly communication, open science, knowledge engineering
- Semantics, Ontology, and Standardisation
- Training, Stewardship, and Data Management Planning

Collaborative Notes: <https://bit.ly/43T1NPd>

Group webpage: [Metadata IG](#) & [Reproducibility IG](#)



BoF: Metadata for Reproducibility



RESEARCH DATA ALLIANCE
24th Plenary Meeting
V I R T U A L

Data for Emerging Technologies

Presenters: Claire Austin, Alex Ball, Lauren Cadwallader,
Milan Ojsteršek, Limor Peer



Agenda

- Welcome & Introductions
- Brief overview of the Metadata and Reproducibility IGs
- Problem statements
- Topics
- Open discussion & wrap up

Speakers

Disclaimer: All views and opinions expressed are those of the co-authors, and do not necessarily reflect the official policy or position of their respective employers, or of any government, agency, or organization.



Limor Peer

イエール大（米）
社会政策研究所
（ISPS）
研究・戦略イニシアチブ担
当アソシエイト・ディレクター

Alex Ball



バース大学（英）
リサーチデータ・ライブラリアン



Lauren
Cadwallader

PLOS
オープンリサーチ
マネージャー

Milan
Ojsteršek



マリボル大学（スロベニア）教員



Claire Austin

Group of Canada
の Science
enablement unit
長



セッション企画IGについて

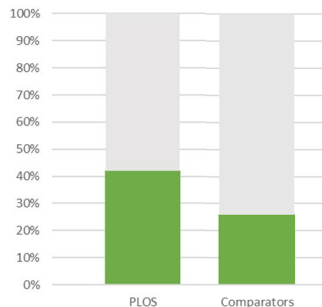
- Metadata IG
研究データに関連するあらゆる側面のメタデータに取り組むことを
目的に活動
- Reproducibility IG
コードを用いるデータや、研究において、再現可能性を促進する
ことを目指して活動



問題提起（社会面）①

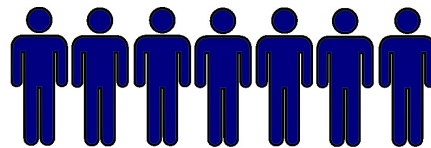
研究者は必ずしも再現可能な形でコードを共有する
必要性を感じていない？

42% of PLOS and 26% of
comparator articles shared
code in 2024

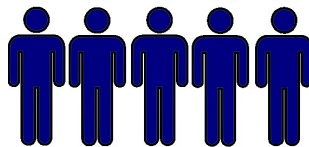


Why do readers look at code?

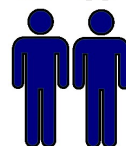
70% aid
understanding
of article



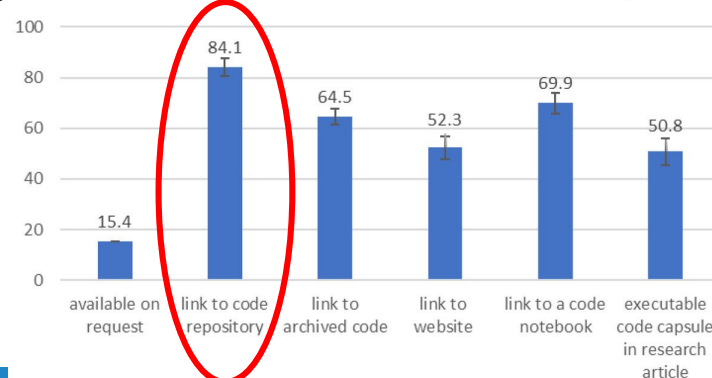
48% reuse



21% research
quality, incl.
reproducibility



Mean usefulness of methods of code sharing



Refs:

Cadwallader & Hrynaskiewicz (2022) doi:
10.7717/peerj.13933

PLOS (2022) doi: 10.6084/m9.figshare.21687686.v9



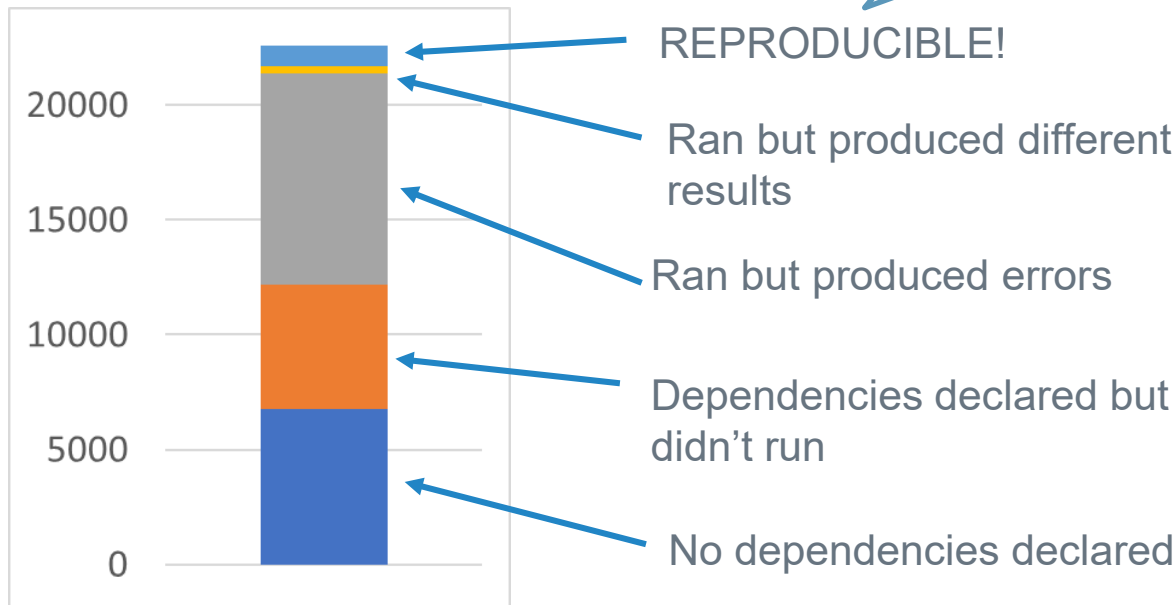
問題提起（社会面）②

研究者は必ずしも再現可能なコードの共有方法を知らない？

Samuel & Mietchen (2024)
DOI: 10.1093/gigascience/giad113

Studied reproducibility of notebooks
written in python in biomedicine

メンバーからの意見：
スキルギャップが原因？
倫理的な問題も？





問題提起（技術面）①課題

- ・ソフトウェアとコードに関するもの（バージョン管理、コードの可用性など）
- ・データ管理と来歴
- ・計算環境
- ・標準化されたメタデータスキーマの欠如
- ・不十分なメタデータ
- ・スキルギャップ

参加者からの意見：

- ・CURE原則、FAIR原則に則ったガイドラインの作成（メタデータ含む）
- ・再現性確保のため要検討
研究者のメリット
再現性の自動評価の仕組み
- ・Baby stepsで進める



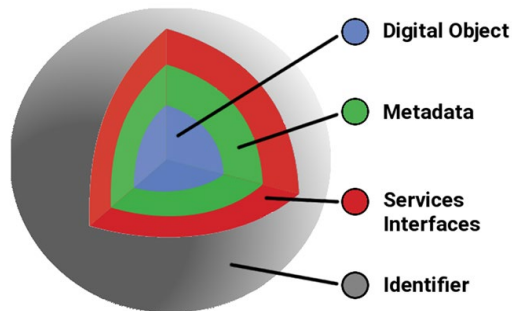
Topic

1. FAIR Digital Object
2. PID graph
3. EOSC minimum metadata set recommendation for metadata interoperability
4. **maFDOs**
machine-actionable FAIR Digital Objects
5. **maDMP**
machine-actionable Data Management Plan
6. 再現性の検証

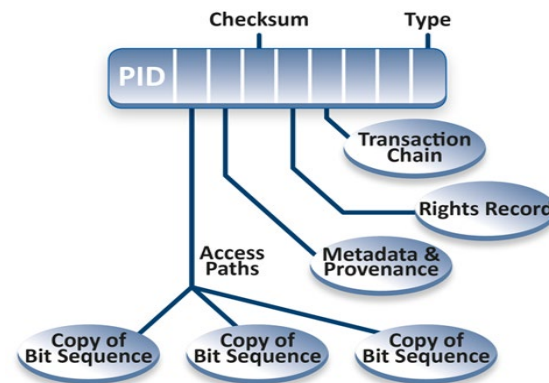
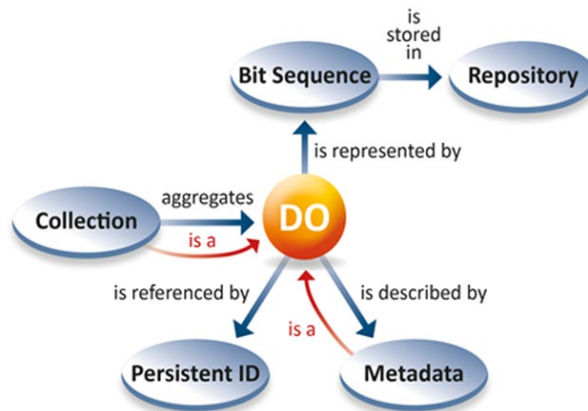


Topic1 : FAIR Digital Object

デジタルオブジェクトと FAIR
原則の統合により、データの
相互運用性と再利用性を向上



Digital Object Interface Protocol

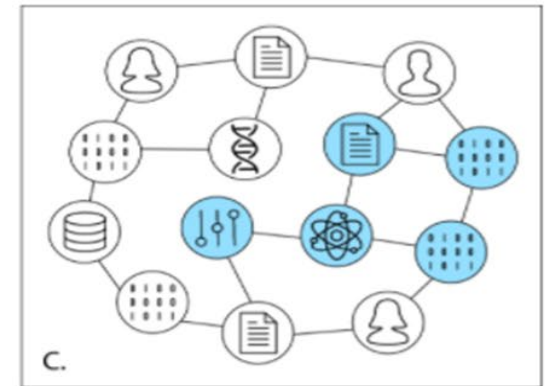
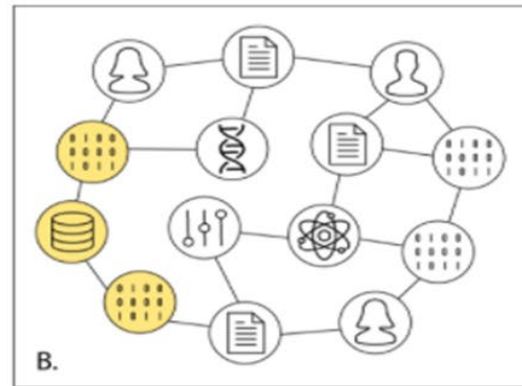
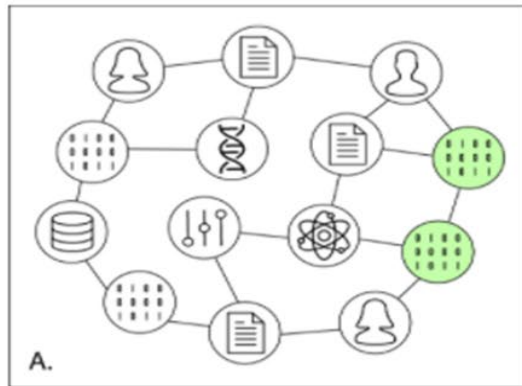


Source: Schwardmann, U., 2020. Digital Objects – FAIR Digital Objects: Which Services Are Required?. *Data Science Journal*, 19(1), p.15. DOI: <http://doi.org/10.5334/dsj-2020-015> and <https://fairdo.org/>



Topic2 : PID graph

永続識別子（Persistent Identifiers, PIDs）とそのメタデータをもとに、論文やデータセット、ソフトウェア、研究者、助成機関などの相互関係を可視化するネットワーク



A – linkage of different version of software

B – datasets stored in the repository

C – linkage of different digital objects of the research project





Topic4 : maFDOs

machine-actionable FAIR Digital Objects

machine-actionable :

機械可読で、機械が解釈可能で、かつ明確に規定された
処理手順に基づいて自動的に実行可能なもの
→ FAIR Digital Objectが目指しているもの

FDO Forum: <https://fairdo.org/>

FDO specifications: <https://fairdo.org/specifications/>

FAIR Digital Object Technical Overview: <https://doi.org/10.15497/rda00031>



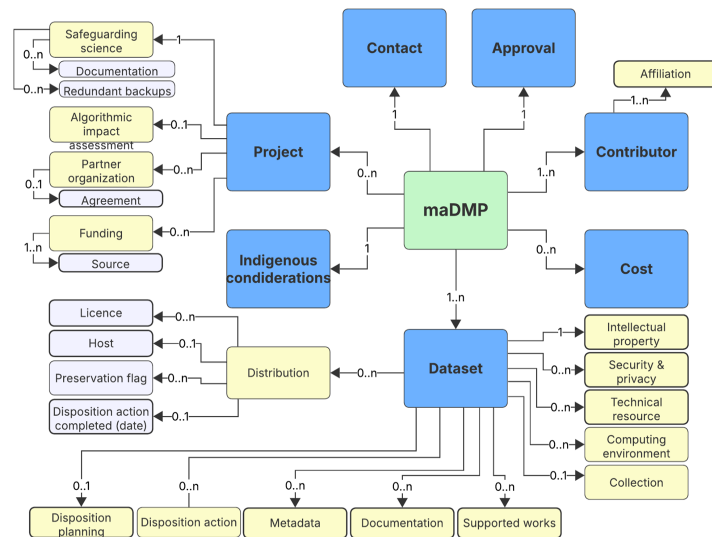
Topic5 : maDMP

machine-actionable Data Management Plan

RDA maDMP Metadata Application Profile (MAP)

- 数年前に、データライフサイクルを通じて継続的にプロジェクトなどの情報を機械的に収集し、DMPを更新する国際標準の仕組みを開発し、現在さらなる拡張を検討中
- FAIRER原則で運用
(Findable, Accessible, Interoperable, Reusable, Ethical, and Reproducible)

GCWG-RDA maDMP Conceptual Model:





Topic6：再現性の検証

「同じデータセットに対して同じ分析ステップを実行すると、一貫して同じ答えが得られるとき、その結果は再現可能である」（Turing Wayの定義）

適用対象をプロジェクトor研究レベルとするとした場合、

- ・誰が検証するのか？
- ・いつ検証するのか？（公開後？）
- ・どのように記載するのか？
- ・検証の本当の意味とは？



Topic6 : 再現性の検証

Certification & badges

Some examples....

IEEE

- **Code available:** The code provided by the authors, including any associated data and documentation, is reasonable and complete and can potentially be used to support the reproducibility of the published results.
- **Code reviewed:** The code provided by the authors, including any associated data and documentation, is reasonable and complete, produces the described outputs when it is run, and can support the reproducibility of the published results.

TOP Guidelines



American Economic Association



Data and Code Availability Standard

ACM



ISPS YALE

Data Quality Review



File Review



Doc Review



Data Review



Code Review

Cornell Center for Social Science



ARTICLE | [Full Access](#)

Why masses support democratic backsliding

Noam Gidron, Yotam Margalit, Lior Sheffer, Itamar Yakir

First published: 23 March 2025 | <https://doi.org/10.1111/ajps.12958>

Yale Links

Verification Materials: The materials required to verify the computational reproducibility of the results, procedures, and analyses in this article are available on the *American Journal of Political Science* Dataverse within the *Harvard Dataverse Network*, at: <https://doi.org/10.7910/DVN/AJCSBJ>

The Cornell Center for Social Sciences verified that the data and replication code submitted to the AJPS Dataverse replicates the numerical results reported in the main text of this article.

Artifact Review and Badging

A variety of research communities have embraced the goal of reproducibility in experimental science. [Learn More](#)



5 Artifacts Available



Artifacts Available

Artifact is supplied in an extremely well packaged manner that enables ease of its reuse.



4 Artifacts Evaluated & Functional



Artifacts Evaluated & Functional

Artifact is supplied in a clear and understandable form.



3 Artifacts Evaluated & Reusable



Artifacts Evaluated & Reusable

Artifact is made available for public distribution.



0 Results Replicated



Results Replicated

Artifact is reviewed using the same experimental setup used by the Author and the results are consistent with those stated by the Author.



7 Results Reproduced



Results Reproduced

A new Artifact is independently developed based on the Author's Artifact and the results are consistent with those stated by the Author.



Topic6 : 再現性の検証

- Cornell Center for Social Sciences



ARTICLE |  Full Access

Why masses support democratic backsliding

Noam Gidron  Yotam Margalit, Lior Sheffer, Itamar Yakir

First published: 23 March 2025 | <https://doi.org/10.1111/ajps.12958>

Verification Materials: The materials required to verify the computational reproducibility of the results, procedures, and analyses in this article are available on the *American Journal of Political Science* Dataverse within the Harvard Dataverse Network, at: <https://doi.org/10.7910/DVN/IXCCBI>.

The Cornell Center for Social Sciences verified that the data and replication code submitted to the AJPS Dataverse replicates the numerical results reported in the main text of this article.



Topic6 : 再現性の検証

- 再現性の評価のためのチェックリストやオンラインツール例
 - FAIRER Aware Reproducibility Checklist
<https://fairerdata.github.io/FAIRER-Aware-REPRODUCIBILITY-Assessment/>
 - TIER2 Pilot 4 Reproducibility Checklist
<https://tier2-project.eu/pilots/4>
 - Reproducibility Checklist (AAAO-Association for the Advancement of Artificial Intelligence)
<https://aaai.org/conference/aaai/aaai-23/reproducibility-checklist/>
 - AI-reproducibility (McGill University and UWaterloo)
<https://www.cs.mcgill.ca/~jpineau/ReproducibilityChecklist.pdf>
 - MICCAI Hackathon
<https://github.com/JunMa11/MICCAI-Reproducibility-Checklist>

FAIRER Reproducibility Checklist

This tool responds to the GC Data Strategy for the Federal Public Service (2023-2026) [Mission 3.3](#) for responsible, transparent and ethical data stewardship to maintain trust. This tool is suitable for a first screening level estimation of the reproducibility of a particular data asset and associated code.

There are two companion tools which you may also find useful:

- [FAIRER-Aware-SELF-Assessment](#) (responds to the GC data strategy Priority 4.2.b., assessment of data skill needs)
- [FAIRER-Aware-DATA-Assessment](#) (responds to the GC data strategy Priority 2.2.b., assessment of existing data for reuse).
- [FAIRER-Aware-AI Safety-Checklist](#) (responds to the GC data strategy Mission 3.3 for responsible, transparent and ethical data stewardship to maintain trust).

Do you work with **data**? Are you looking to make it future proof? The **FAIRER data principles** will help you!

In 2023, the international consortium, [Common Infrastructure for National Cohorts in Europe, Canada, and Africa \(CINECA\)](#) stated that: "While the FAIR principles have become a guiding technical resource for data sharing, legal and socio-ethical considerations are equally important for a fair data ecosystem. ... FAIR data should be FAIRER, including also ethical and reproducible as key components."

FAIRER principles refer to the **Findability, Accessibility, Interoperability, Reusability, Ethics and Reproducibility** of data assets, including related code. Applying these principles to your data assets will help others to find, verify, cite, and reuse your data and code more easily.

This tool helps you to assess the reproducibility of a data asset and associated code, and get tips on how you could increase their value and impact.

The tool is discipline-agnostic, making it relevant to any scientific field.

The checklist will take 15-30 minutes to complete, after which you will receive a quantitative summary of the level of reproducibility of your data and code, and tips on how you can improve the level of reproducibility. No information is saved on our servers, but you will be able to save the results of the assessment, including tips for improvement, to your local computer and add notes for future reference.

[CRedit](#) [Author statement](#)

REPRODUCIBILITY:

Reproducible data and code means that the final data and code are computationally reproducible within some tolerance interval or defined limits of precision and accuracy, i.e. a 3rd party will be able to verify the data lineage and processing, reanalyze the data and obtain consistent computational results using the same input raw data, computational steps, methods, computer software & code, and conditions of analysis in order to determine if the same result emerges from the reprocessing and reanalysis. "Same result" can mean different things in different contexts: identical measures in a fully deterministic context, the same numeric results but differing in some irrelevant detail, statistically similar results in a non-deterministic context, or validation of a hypothesis. All data and code are made available for 3rd-party verification of reproducibility. Note that reproducibility is a different concept from replicability. In the latter case, the final published data are linked to sufficiently detailed methods and information for a 3rd-party to be able to verify the results based on the independent collection of new raw data using similar or different methods but leading to comparable results.

Checklist Questions

My code is:

1 Deterministic	☐ Yes ☐ No
1 Non-Deterministic	☐ Yes ☐ No
Executable only on a high performance computing (HPC) system or a super computer.	☐ Yes ☐ No
1 Quantum Code	☐ Yes ☐ No

For all models and algorithms, I provided a link to:

A clear description of the mathematical setting, algorithm, and/or model.	☐ Yes ☐ No ☐ Partial ☐ N/A
A clear explanation of any assumptions.	☐ Yes ☐ No ☐ Partial ☐ N/A
An analysis of the complexity (time, space, sample size) of the algorithm.	☐ Yes ☐ No ☐ Partial ☐ N/A
A conceptual outline and/or pseudocode description.	☐ Yes ☐ No ☐ Partial ☐ N/A

For any theoretical claims, I provided a link to:

A clear statement of the claim.	☐ Yes ☐ No ☐ Partial ☐ N/A
A complete proof of the claim.	☐ Yes ☐ No ☐ Partial ☐ N/A
A clear formal statement of all assumptions.	☐ Yes ☐ No ☐ Partial ☐ N/A
A clear formal statement of all restrictions.	☐ Yes ☐ No ☐ Partial ☐ N/A
Proofs of all novel claims.	☐ Yes ☐ No ☐ Partial ☐ N/A

<https://fairerdata.github.io/FAIRER-Aware-REPRODUCIBILITY-Assessment/>



ディスカッション

- 誰が研究の再現性について検証すべきか？
→ 内部検証の仕組みがあると良いのでは
- 研究者が再現性の担保が自分たちのためのものであると認識すれば、
検証しやすい仕組みにデータを整備するようになる
- 出版社が再現性について方針を示す等していることは望ましいこと
- 学部や大学院のプログラムにも再現性に関する内容を組み込むべき
- サンプルデータをリクエストするのは？ 機密データにも適用できるか？
→ 制限があるデータも、研究者が再現目的で許可を得れば、取得可能
(でも難しそう)



再現性確保に向けた提案

研究者が自分のコードの再現性を検証できるように、まずすべきこととして、

- ・アドバイスや推奨事項の提示
- ・研究者が使用可能な技術やアプローチの提示
- ・推奨事項に関するトレーニング

生成AIの利用や
学際的研究の推進
などから、日本でも
今後課題に!?