



プレプリントをめぐる動向

—COVID-19が加速する学術情報流通の変容—

林 和弘

文部科学省科学技術・学術政策研究所

データ解析政策研究室長

第5回学術コミュニケーションセミナー（月刊JPCOAR）

2022年2月21日（月）14時～16時

学術情報流通のオープン化をめぐる諸問題

1. はじめに

- ✓ 学術情報流通の変革とオープンサイエンス

2. プレプリントとプレプリントサーバーについて

- ✓ プレプリントは今も昔も
- ✓ プレプリントサーバーの進展

3. COVID-19とプレプリント

- ✓ COVID-19で医学系にも浸透し始めたプレプリント
- ✓ プレプリントがもたらす価値と可能性

4. プレプリントの展望

- ✓ プレプリントとの向き合い方
- ✓ オープンサイエンスに向けて

5. まとめ

1. はじめに

- ✓ 学術情報流通の変革とオープンサイエンス

2. プレプリントとプレプリントサーバーについて

- ✓ プレプリントは今も昔も
- ✓ プレプリントサーバーの進展

3. COVID-19とプレプリント

- ✓ COVID-19で医学系にも浸透し始めたプレプリント
- ✓ プレプリントがもたらす価値と可能性

4. プレプリントの展望

- ✓ プレプリントとの向き合い方
- ✓ オープンサイエンスに向けて

5. まとめ

■ 1990年代:デジタル化とweb化

- ◆ 東京大学・理学部(有機合成化学))で研究しながら、学術情報流通の変革に実地で参画(化学者 & ITオタクとして)
- ◆ 試薬管理DB、ジャーナル査読DBから電子ジャーナル開発へ



■ 2000年代:web活用の本格化とオープン化

- ◆ 日本化学会(常勤で15年)にて、電子投稿査読開発、電子ジャーナル化、世界最速レベルの出版体制構築、ビジネスモデルの確立、オープンアクセス対応などをこなす(出版者として)

■ 2010年代:研究成果の拡張(論文からデータ)と研究インパクト計量の新展開



- ◆ 科学技術・学術政策研究所で(現在9年目)、科学技術予測調査の傍ら、altmetrics等新しい研究評価手法の調査、オープンアクセス、オープンサイエンス政策などの調査研究と実装に取り組む(内閣府、RDA、OECD、G7)

■ 2020年代:オープンサイエンスと科学の変容

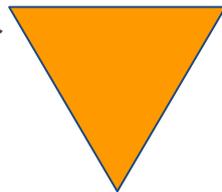


1990年代よりICTを活用した”科学の社会”問題解決を志向&試行し, 多様なステークホルダーに自ら飛び込んでオープンサイエンスパラダイムへの変容(DX)を促す触媒型研究者

政策

専門委員他として
ガイドライン・ポリシー
作成等に関わる

- UNESCO
- G7科技大臣会合
- OECD
- 内閣府
- 文科省



現場

アカデミア

分野を超えた対話の繰り返しと啓発

- 日本学術会議特任連携会員(オープンサイエンス他)
- 千葉大学非常勤講師(学術情報論)
- 京都大学アカデミックデータ・イノベーションユニットメンバー
- 複数の学会・学術雑誌の編集委員, アドバイザー等



セクターを超え, 実践に基づく対話の繰り返しと啓発

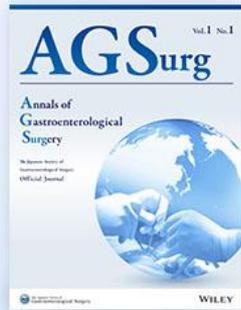
- 有機合成化学専攻(東大:DC1を取ったが途中で方針変更)
- 黎明期の電子ジャーナル開発と学会運営(日本化学会, J-STAGE)
- 大学図書館との未来洞察(SPARC Japan)
- 学術情報流通の啓発(OA, altmetrics, プレプリント, ORCID, PID→定量的研究評価の理想と現実)
- 研究データ利活用の実践と啓発(RDA, 研究データ利活用協議会)



知見を転用して
PTAの電子化も
無理なくサクッと

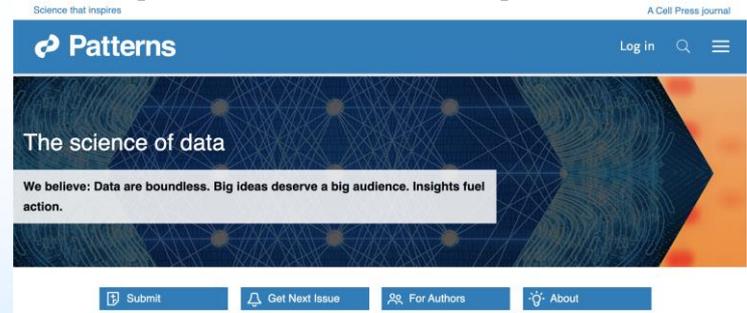
日本消化器外科学会 英文誌アドバイザー

Cell Press, Patterns (Science of Data)



Annals of
Gastroenterological
Surgery

<https://onlinelibrary.wiley.com/journal/24750328>



<https://www.cell.com/patterns/home>

日本医学雑誌編集者会議 組織委員会委員 (JAMJE)



日本医学雑誌編集者会議 Japanese Association of Medical Journal Editors (JAMJE)

組織

日本医学雑誌編集者組織委員会
(任期：2019.10.1～2021.6 日本医学会臨時評議員会開催日迄)

○...委員長

担当副会長

岸 玲子 北海道大学環境健康科学研究教育センター特別招へい教授/名誉教授

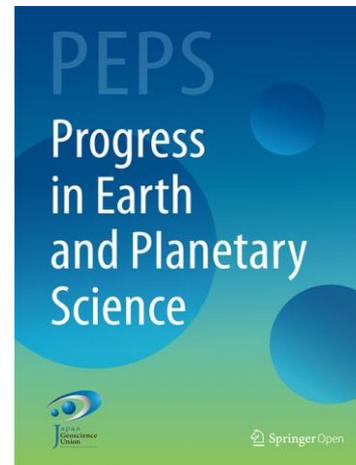
委員

有馬 寛 名古屋大学大学院 医学系研究科 教授

遠藤 格 横浜市立大学大学院 医学研究科 教授

<http://jams.med.or.jp/jamje/iinkai.html>

日本地球惑星科学連合



<http://progearthplanetsci.org/>

国際純正・応用化学連合 (IUPAC) ~2019



<https://iupac.org/>

I U P A C

科学論文を書き、編集・出版・電子化・OA化に主体的に関わり、プレプリントも公開した経験と知見を共有

1. はじめに

- ✓ 学術情報流通の変革とオープンサイエンス

2. プレプリントとプレプリントサーバーについて

- ✓ プレプリントは今も昔も
- ✓ プレプリントサーバーの進展

3. COVID-19とプレプリント

- ✓ COVID-19で医学系にも浸透し始めたプレプリント
- ✓ プレプリントがもたらす価値と可能性

4. プレプリントの展望

- ✓ プレプリントとの向き合い方
- ✓ オープンサイエンスに向けて

5. まとめ

- **プレプリント = 査読つきジャーナルに通る前の草稿原稿**
 - ◆ プレプリントは査読済みでもなく、出版（Publish）されたものでもない
 - ◆ 紙の時代からあるプレプリントの共有（身内，共同研究者）
- **プレプリントサーバー = プレプリントを掲載するインターネットサーバー（Web）**
 - ◆ 1991年から30年あまりの歴史（物理）
 - ◆ “身内と共有”から“世界と共有”へ

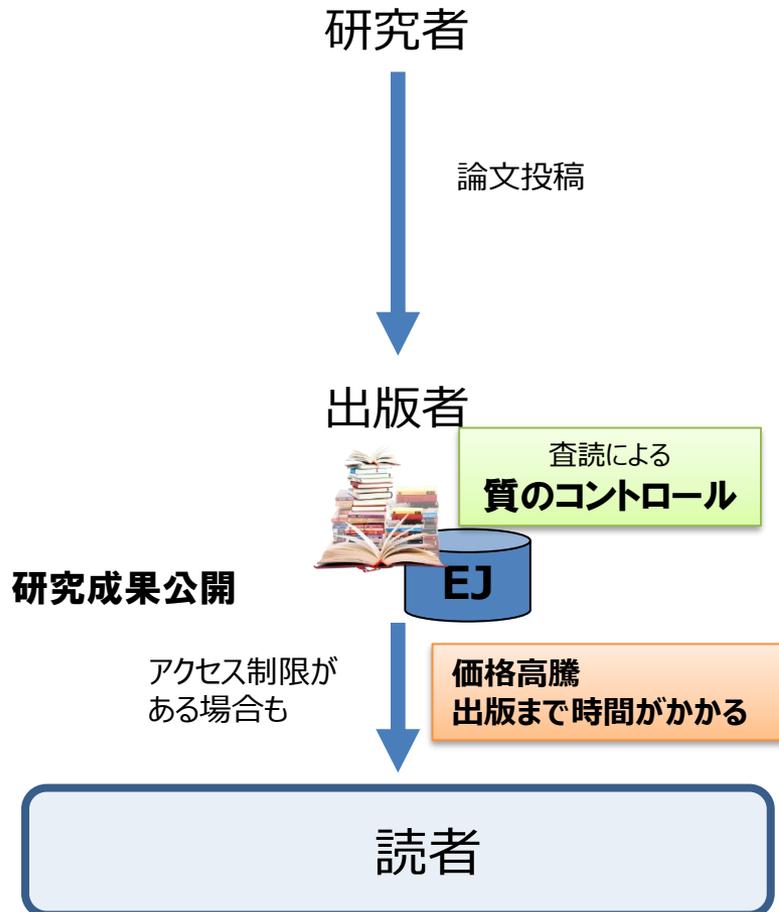


<https://www.medrxiv.org/>

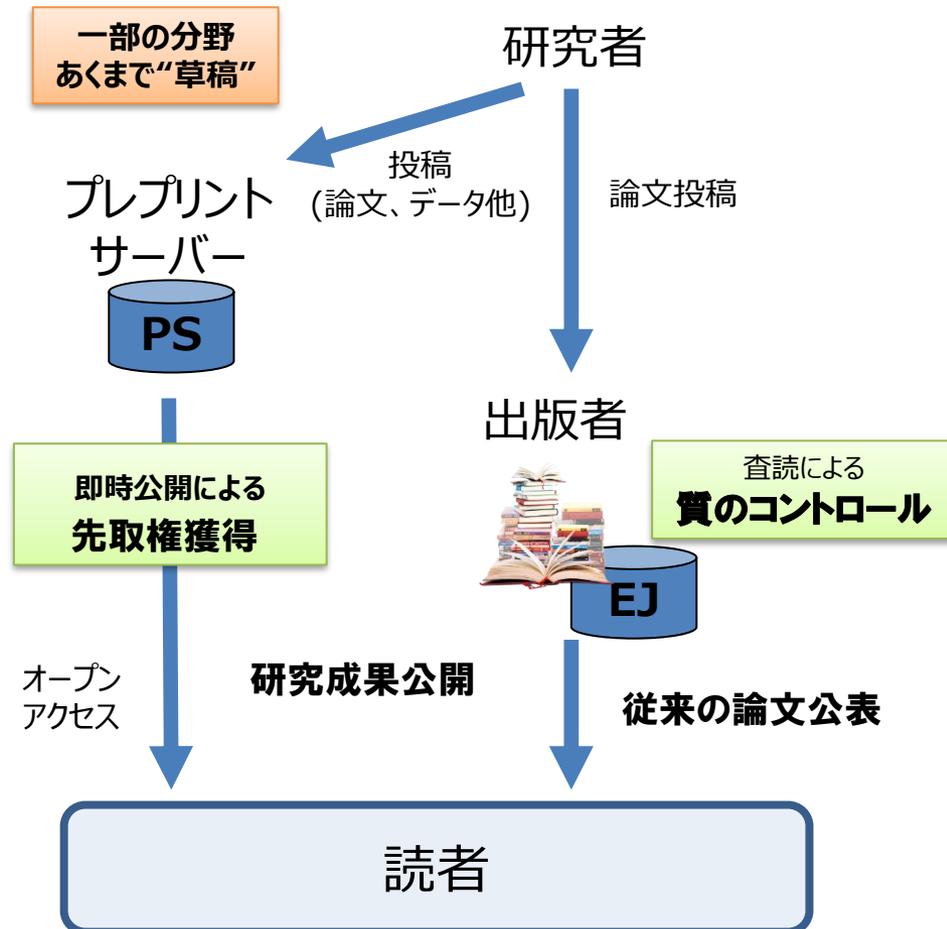
2019年6月5日発足
米イェール大学、BioRxivを立ち上げたコー
ルドスプリングハーバー研究所にBMJ(British
Medical Journal)が参画して始動

- **Caution: Preprints are preliminary reports of work that have not been peer-reviewed. They should not be relied on to guide clinical practice or health-related behavior and should not be reported in news media as established information.**

従来の仕組み



プレプリントサーバ (PS) の活用

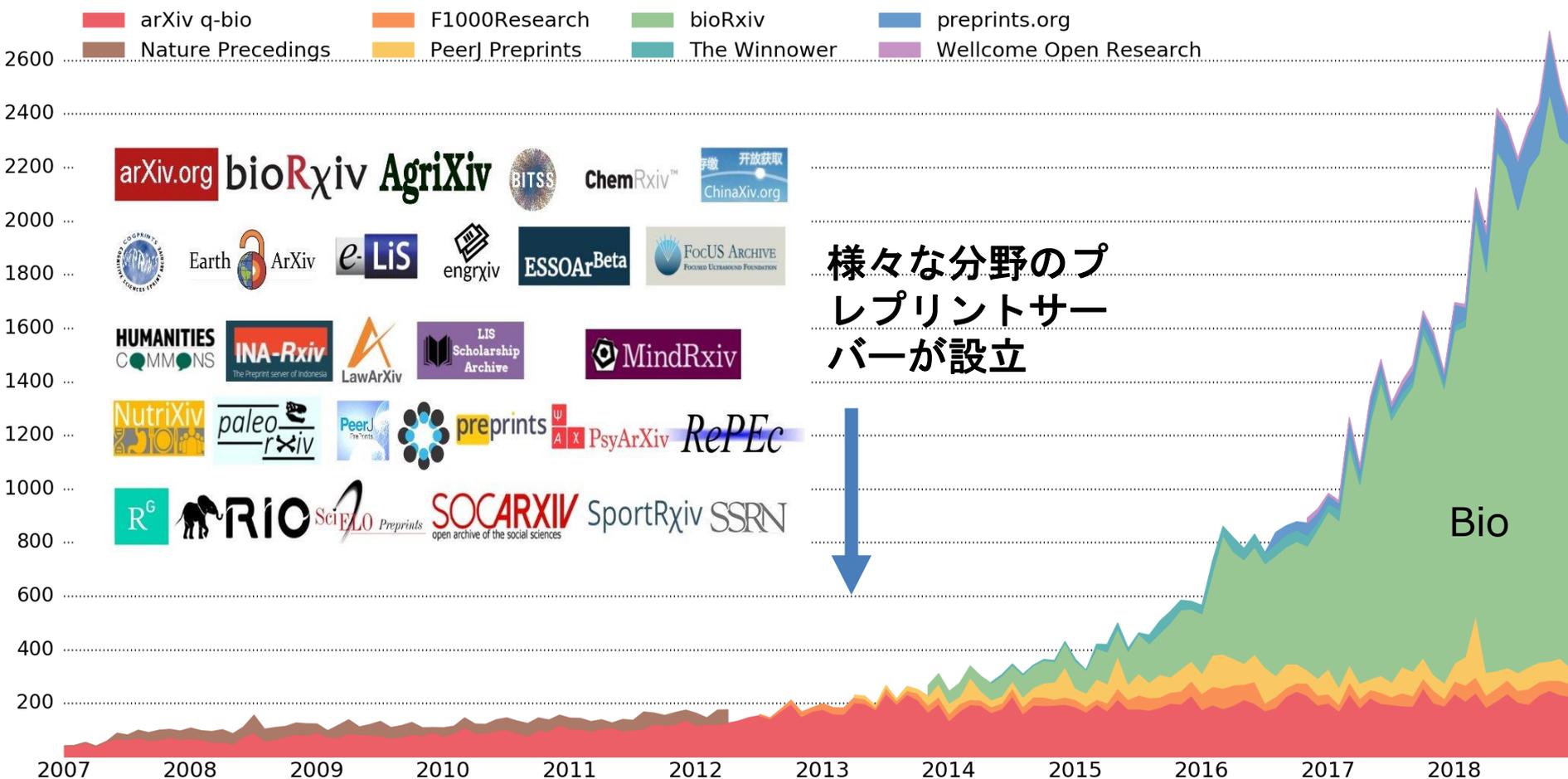


先取権を確保しつつ、査読を経て学術ジャーナルでも公開
領域によってはプレプリントサーバがメインの公開先になるところも

名称	創設年	2020年1月現在の 運営母体	分野	システム	DOI
arXiv	1991	コーネル大学	物理学に始まり, 情報学, 経済学等 多分野に広がる	オリジナル	×
SSRN	1994	Elsevier	社会科学に始まり多分野に広がる	オリジナル (ColdFusion)	○
BioRxiv	2013	コールド・スプリング・ハーバー研究所	生命科学を 中心とした分野	HighWirePress	○
ChemRxiv	2017	米国化学会, 英国化学会, ドイツ化学 会, 日本化学会, 中国化学会	化学を 中心とした分野	figshare	○
MedRxiv	2019	米イェール大学, コールド・スプリングハーバー研究所, BMJ(British Medical Journal)	医学を 中心とした分野	HighWirePress	○

迅速な成果公開・共有手段として分野ごとに徐々に浸透中
学術情報流通を変革する研究データより身近なメディアとして注目

Preprints per Month

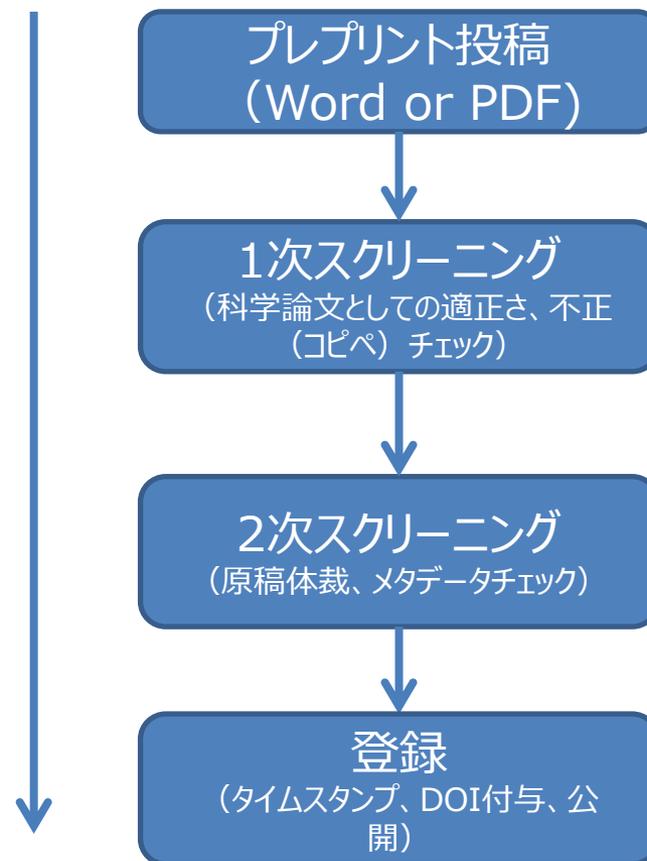


arXive (高エネルギー物理→統計、数学) →生物→AI

http://www.prepubmed.org/monthly_stats/

■ 科学論文としての最低限の チェックが行われる

- ◆ どのプレプリントサーバーもいわゆる“トンデモ論文”は事前にフィルターしている
- ◆ コピペチェックまでしているところは少ない
- ◆ 中身の新規性、速報性には触れない



4-48hours

1. はじめに

- ✓ 学術情報流通の変革とオープンサイエンス

2. プレプリントとプレプリントサーバーについて

- ✓ プレプリントは今も昔も
- ✓ プレプリントサーバーの進展

3. COVID-19とプレプリント

- ✓ COVID-19で医学系にも浸透し始めたプレプリント
- ✓ プレプリントがもたらす価値と可能性

4. プレプリントの展望

- ✓ プレプリントとの向き合い方
- ✓ オープンサイエンスに向けて

5. まとめ

- **1969年インゲルフィンガー・ルール (Ingelfinger Rule)**
 - ◆ その内容が既に投稿されていたり、どこか別の場所で報告されていたりした場合は、原稿の受け取りを拒否する
- **1999年E-Biomed by H. Varmusが理論物理学を真似る形で医学版arXivを提唱するも頓挫**
 - ◆ 内容を理解できる人が限られる理論物理学とは異なり、健康や生命に直接的にかかわり多くの人々が興味を持つ医学医療分野の懸念
 - ◆ これを機にPMC（査読済み論文のポータル）創立、PLoSジャーナル（Gold OAジャーナル）創刊へ
- **2003年arXivが定量生物学の論文を受付**
 - **2013年BioRxivの誕生と急進**
 - **2017年MedRxivを提唱**
 - **2019年に始動**

■ プレプリントは、査読つきジャーナルを置き換えるものではなく、サービスを補完するもの

- ◆ Webの特性を生かし、新しい知見の迅速な共有とフィードバックを得る
- ◆ 先取権の確保
- ◆ 査読者のScoop（抜け駆け）防止



<https://www.medrxiv.org/>

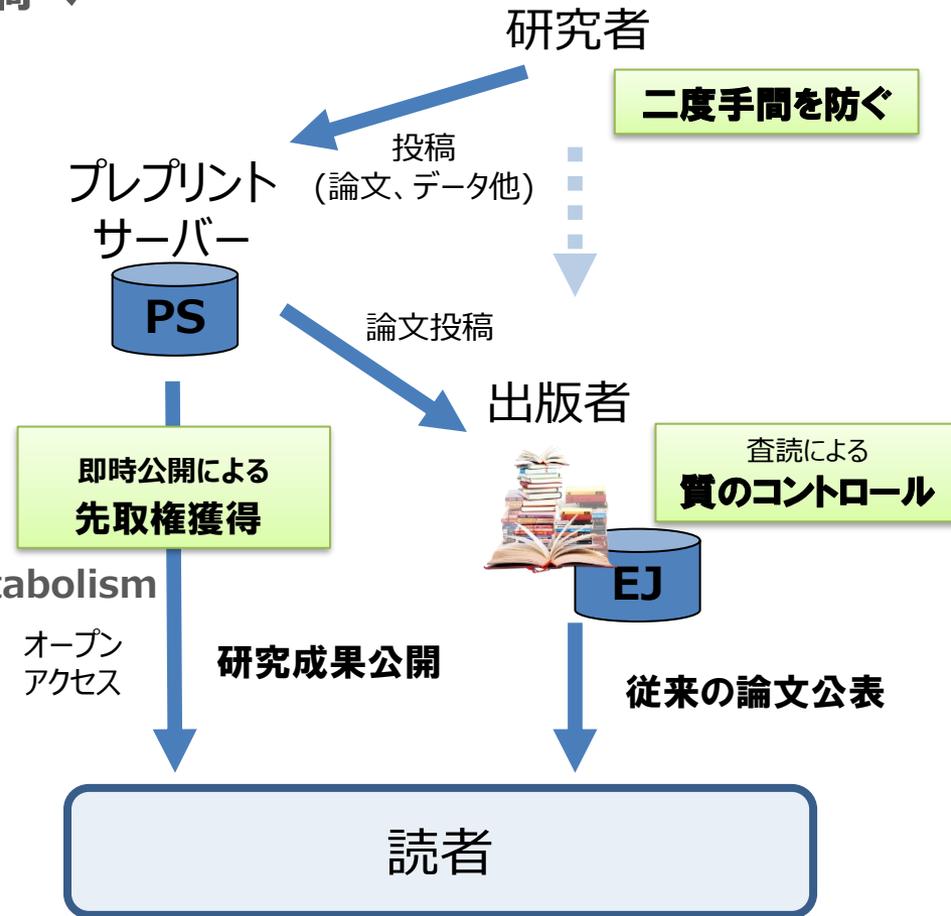
2019年6月5日発足
 米イェール大学、BioRxivを立ち上げたコー
 ルドスプリングハーバー研究所にBMJ(British
 Medical Journal)が参画して始動

- **Caution: Preprints are preliminary reports of work that have not been peer-reviewed. They should not be relied on to guide clinical practice or health-related behavior and should not be reported in news media as established information.**

- Addiction Medicine
- Allergy and Immunology
- Anesthesia
- Cardiovascular Medicine
- Dentistry and Oral Medicine
- Dermatology
- Emergency Medicine
- Endocrinology (including Diabetes Mellitus and Metabolic Disease)
- Epidemiology
- Forensic Medicine
- Gastroenterology
- Genetic and Genomic Medicine
- Geriatric Medicine
- Health Economics
- Health Informatics
- Health Policy
- Health Systems and Quality Improvement
- Hematology
- HIV/AIDS
- Infectious Diseases (except HIV/AIDS)
- Intensive Care and Critical Care Medicine
- Medical Education
- Medical Ethics
- Nephrology
- Neurology
- Nursing
- Nutrition
- Obstetrics and Gynecology
- Occupational and Environmental Health
- Oncology
- Ophthalmology
- Orthopedics
- Otolaryngology
- Pain Medicine
- Palliative Medicine
- Pathology
- Pediatrics
- Pharmacology and Therapeutics
- Primary Care Research
- Psychiatry and Clinical Psychology
- Public and Global Health
- Radiology and Imaging
- Rehabilitation Medicine and Physical Therapy
- Respiratory Medicine
- Rheumatology
- Sexual and Reproductive Health
- Sports Medicine
- Surgery
- Toxicology
- Transplantation
- Urology

プレプリント登録からそのままジャーナル投稿へ

- BMJ Open
- Cells Tissues Organs
- Developmental Neuroscience
- EMBO Molecular Medicine
- Endocrinology
- Genetics in Medicine
- Human Heredity
- Intervirology
- JCO Clinical Cancer Informatics
- JCO Precision Oncology
- The Journal of Clinical Endocrinology & Metabolism
- Journal of the Endocrine Society
- Journal of Experimental Medicine
- Journal of Innate Immunity
- Life Science Alliance
- Microbial Physiology
- PLOS Medicine
- PLOS Neglected Tropical Diseases
- PLOS One
- StemJournal

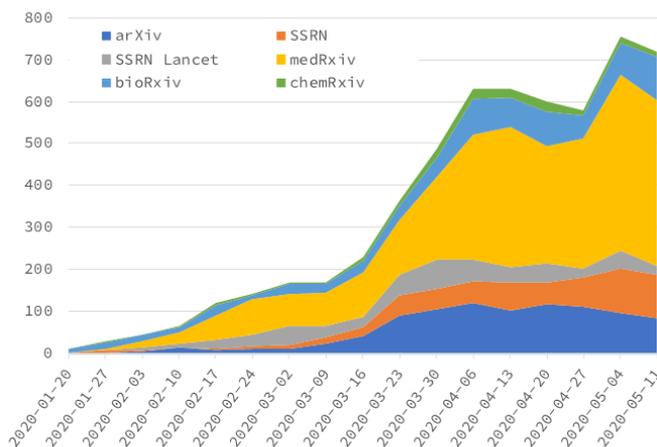
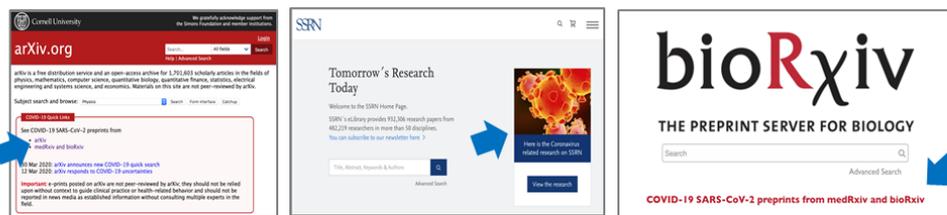


- 投稿者 : 1回の投稿作業で、プレプリントと査読付きジャーナルへの作業が完了
- 出版社 : ある程度のスクリーニングが済んだ原稿を査読できる

■ 原著論文、被引用数を使わずに研究動向の把握(ネットワーク分析)が可能に

■ 多くの PPS が COVID-19 関連論文リストを作成

◆ これらのリストを対象として, COVID-19関連の動向を調査



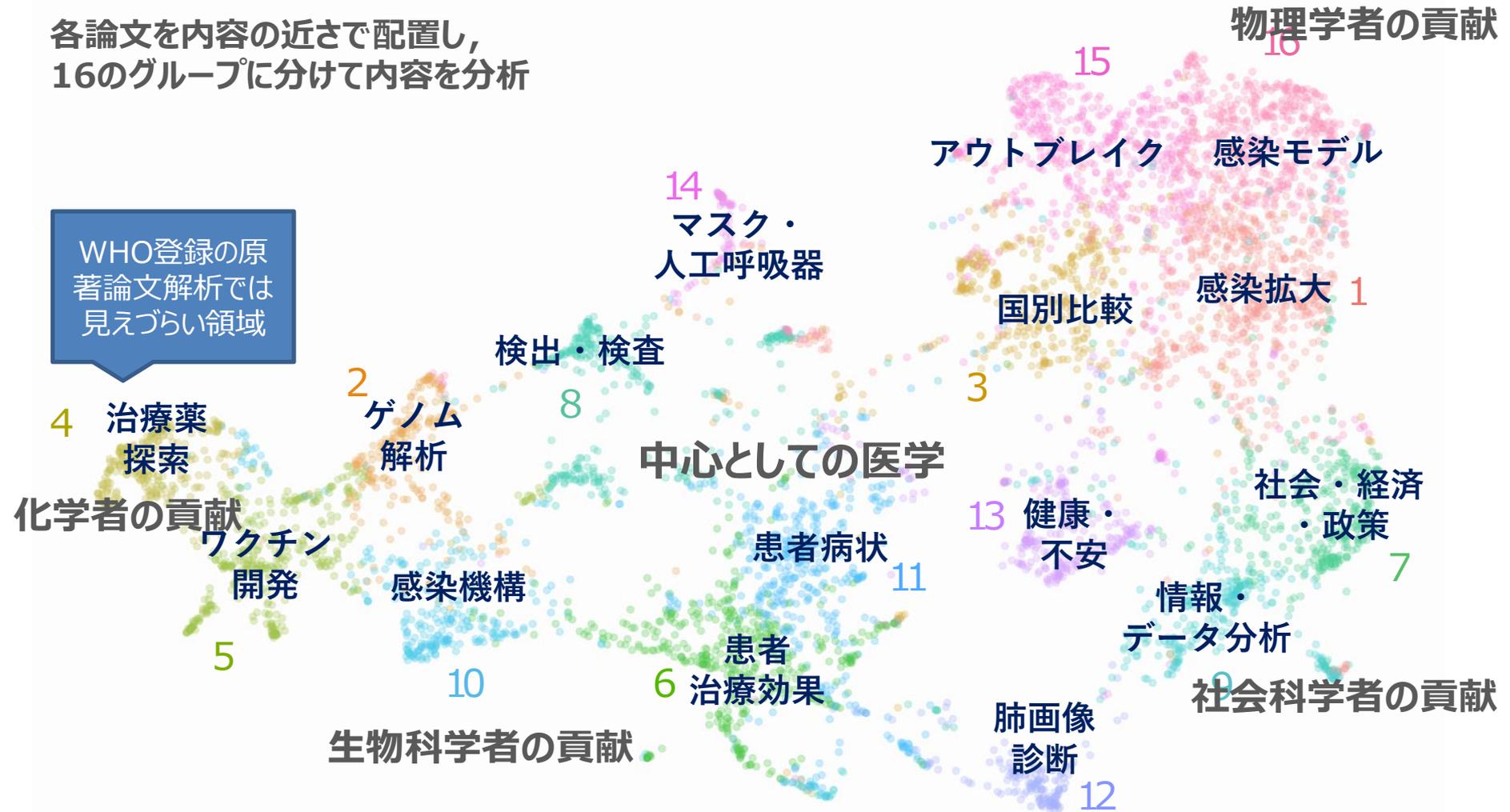
Week	arXiv	SSRN Lancet	medRxiv	bioRxiv	chemRxiv	
2020-01-20 (04)	0	1	0	0	9	1
2020-01-27 (05)	2	5	1	3	16	1
2020-02-03 (06)	4	5	5	16	13	1
2020-02-10 (07)	13	2	8	27	11	5
2020-02-17 (08)	7	4	20	59	24	6
2020-02-24 (09)	10	6	28	84	9	3
2020-03-02 (10)	10	10	45	76	24	5
2020-03-09 (11)	24	13	27	81	22	2
2020-03-16 (12)	42	20	26	106	27	10
2020-03-23 (13)	91	46	49	133	31	13
2020-03-30 (14)	105	50	68	198	47	19
2020-04-06 (15)	120	51	51	300	85	25
2020-04-13 (16)	102	68	35	334	70	21
2020-04-20 (17)	116	52	45	280	84	24
2020-04-27 (18)	111	69	23	309	54	13
2020-05-04 (19)	95	107	43	420	76	14
2020-05-11 (20)	84	103	22	394	104	12

コロナ禍にあって, 医学・生物・化学系だけでなく, 情報, 人社系など多くの分野でプレプリントが活発に投稿されている

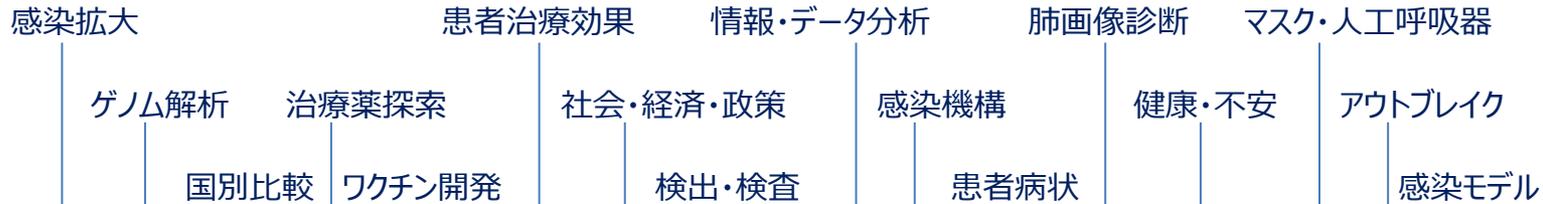
2020.05.18にデータを収集し, 2020年第20週 05/17 までのデータを取得
(各PPSともに記事のPosted Dateを基準として採用)

各論文を内容の近さで配置し、
16のグループに分けて内容を分析

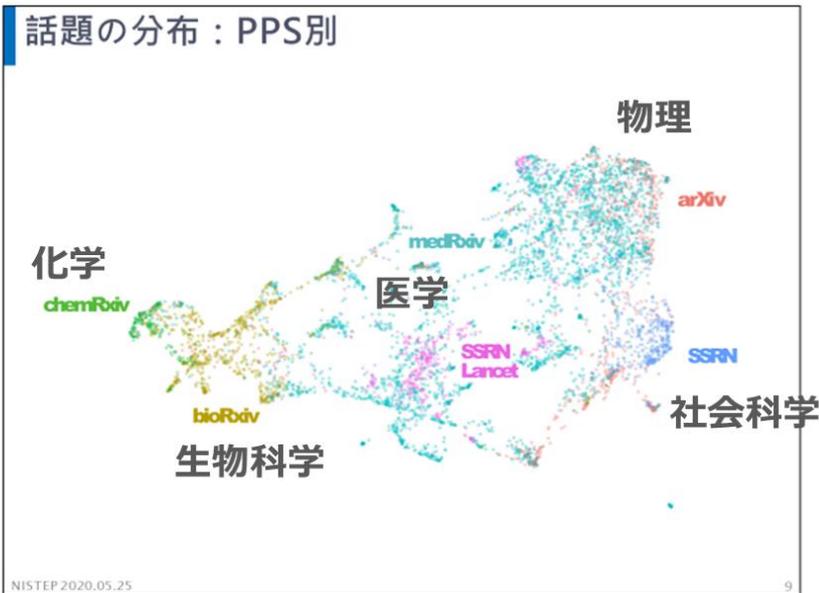
WHO登録の原著論文解析では見えずらい領域



注)原著論文, 被引用数による解析を代替するものではなく、付加的なものとして使い分ける



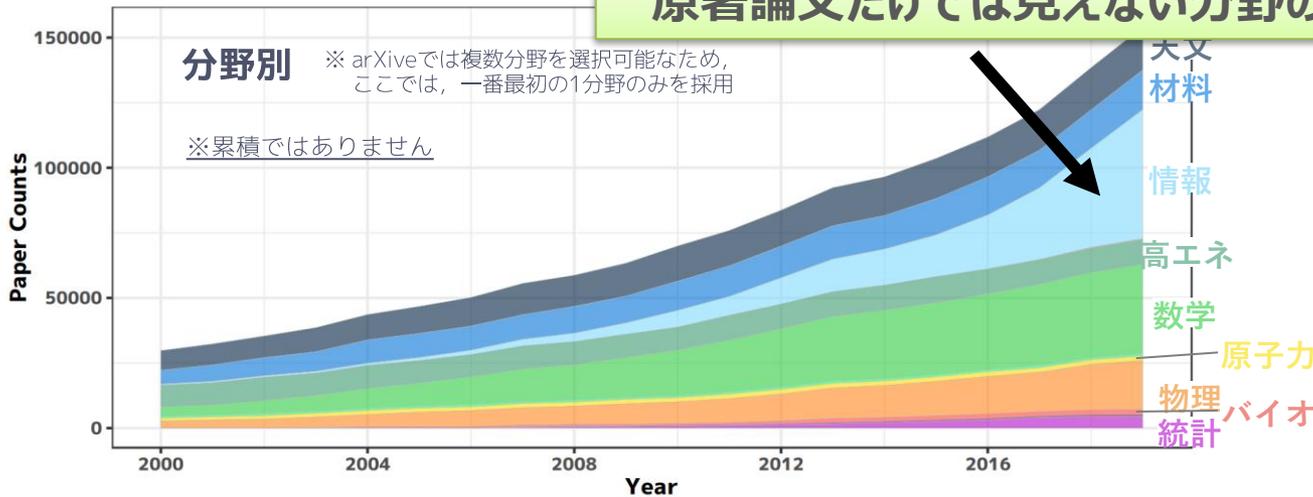
Kind	Total	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
arXiv	936	16.5%	1.3%	3.3%	2.8%	3.0%	0.0%	5.0%	1.6%	19.0%	1.6%	0.0%	10.5%	0.7%	1.8%	7.9%	25.0%
bioRxiv	716	0.1%	19.6%	0.0%	8.9%	40.2%	0.0%	0.0%	9.8%	2.1%	17.5%	0.0%	0.3%	0.1%	0.4%	1.0%	0.0%
chemRxiv	175	0.0%	1.1%	0.6%	75.4%	17.1%	0.0%	0.0%	1.1%	1.1%	1.7%	0.6%	0.0%	0.0%	1.1%	0.0%	0.0%
medRxiv	2837	15.2%	2.3%	11.3%	0.7%	0.5%	13.5%	0.6%	7.8%	3.3%	2.9%	6.6%	2.7%	6.7%	2.7%	13.6%	9.8%
SSRN	612	10.3%	2.5%	7.7%	2.0%	1.5%	2.0%	49.3%	0.8%	7.4%	2.1%	1.8%	0.0%	2.0%	1.3%	6.0%	3.4%
SSRN Lancet	496	2.6%	0.2%	1.6%	0.6%	0.0%	12.9%	1.2%	2.0%	1.4%	2.0%	48.4%	0.0%	12.1%	0.0%	13.9%	1.0%



- 他分野のプレプリントによる迅速な研究成果共有と社会課題解決
 - ◆ 研究データも共有
- プレプリントによる先取権の確保
→“プレプリント文化の進展の可能性”
- ただし、質保証の仕組みが必要（既存の時間がかかる査読形式ではない形で）

- 1991年より物理から始まり、昨今では、AI関連の投稿も多く、物理・数学・情報系で著名
 - ◆ 国際会議なども重視され、論文だけでは動向を追いつらいとされる情報系に有用

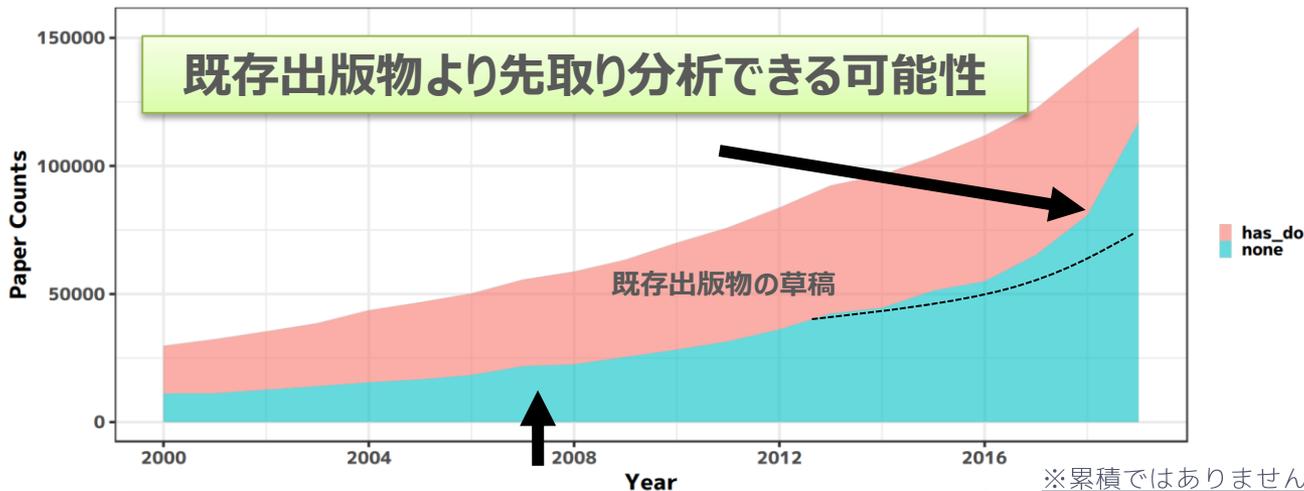
原著論文だけでは見えない分野の動向把握



2019年には年間投稿数が約15万件、累積で160万件を超える

情報系の躍進（登録数の増大+プレプリントを引用する傾向が他より強い）

既存出版物より先取り分析できる可能性



出版年ごとにDOIの有無（≒査読付きジャーナル等の既存の出版物に出ているもの）を見るとプレプリントならではの分析の可能性が見える

（点線は主観による参考）

出版バイアスで除外された知見の可能性

◆ 情報系で定性的に言われていたことをある程度定量的に裏打ち

- 原著論文ではなく，国際会議のプロシーディングスを重要視
- プレプリントを参考に研究を進展させ，プレプリントを引用してプレプリントを共有（Deep Learning）

◆ COVID-19の前より，査読ジャーナルの問題は顕在化しており，分野依存ながら代替手段も存在していた（研究サイクルの速さにジャーナル査読が追いつかない）

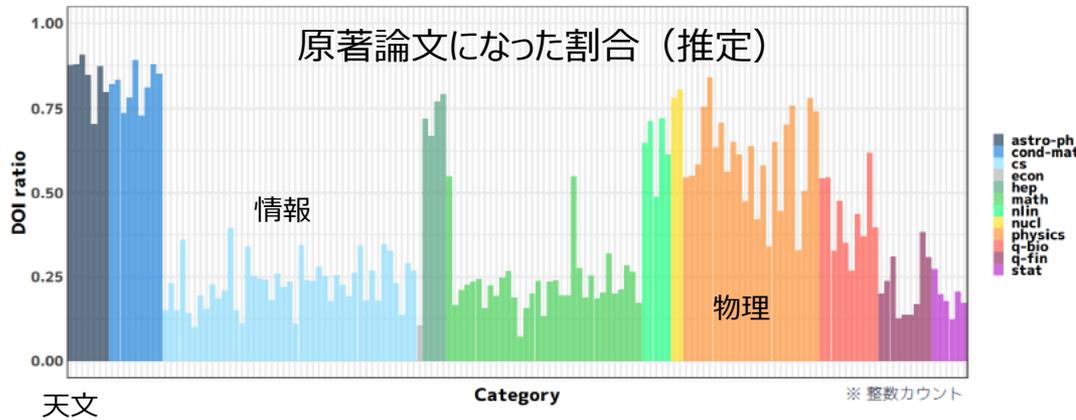


図5 分野毎のDOI付与率

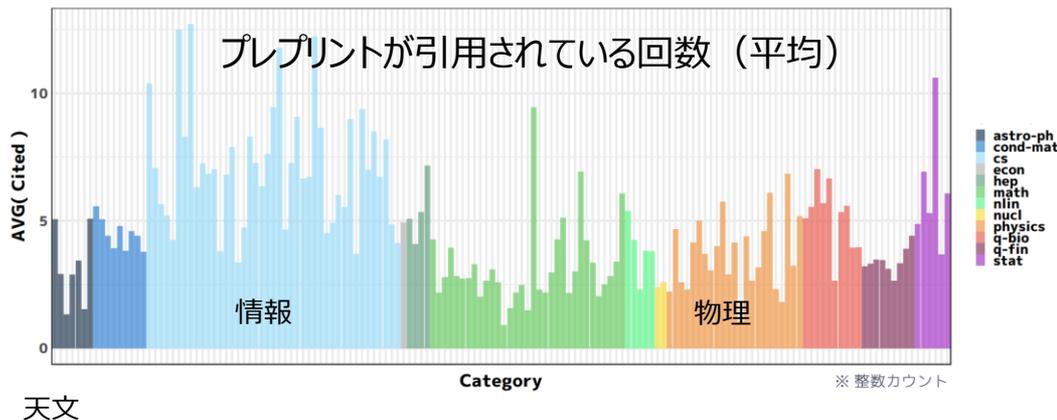


図7 分野と被引用の関係性

astro-ph	天文
cond-mat	材料
cs	情報
econ	計量経済
hep	高エネルギー物理
math	数学
nlin	システム科学
nucl	核
physics	物理
q-bio	生物
q-fin	金融
stat	統計

arXivが設定した153の小分野を独自に12分類

■ 世界的な社会課題 (Global Urgent Issue)

- ◆ COVID-19
 - ◆ ジカ熱
 - ◆ SARS
- 緊急対応

■ 実用性のある最先端科学トピック (Super Hot Topic)

- ◆ Deep Learning
- ◆ 高温超伝導

研究の進展が早すぎて査読が間に合わない (待てない)

■ 新奇性として先進的なトピック (Super Novel Topic)

- ◆ ポアンカレ予想, ABC予想
- ◆ 先進的過ぎると査読に通らない (価値判断ができない、時間がかかりすぎる)

■ 科学的より社会インパクトが強いもの (Super Societal Impact)

- ◆ Google Page Rank、経済系のワーキングペーパー (プレプリント相当)
- ◆ 実用性が高いものを手早く共有

**査読済論文の草稿としてのプレプリントから
査読済み論文とは違う価値を持つオンラインメディアへ**

1. はじめに

- ✓ 学術情報流通の変革とオープンサイエンス

2. プレプリントとプレプリントサーバーについて

- ✓ プレプリントは今も昔も
- ✓ プレプリントサーバーの進展

3. COVID-19とプレプリント

- ✓ COVID-19で医学系にも浸透し始めたプレプリント
- ✓ プレプリントがもたらす価値と可能性

4. プレプリントの展望

- ✓ プレプリントとの向き合い方
- ✓ オープンサイエンスに向けて

5. まとめ

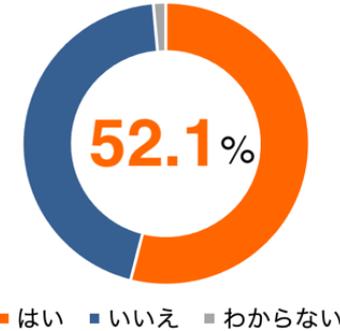
■ 調査対象

- 科学技術専門家ネットワーク：1,914名
- 有効回答：1,448名（回答率75.7%）

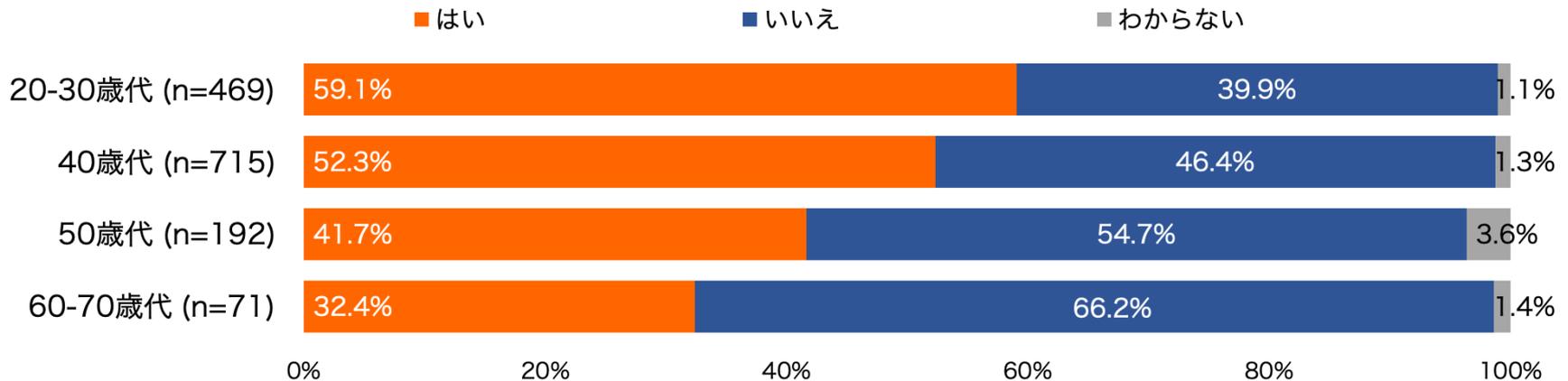
■ 調査期間

- 2020年8月17日～8月31日（9月6日まで）

プレプリントの入手経験（全体）



年齢層別プレプリントの入手経験

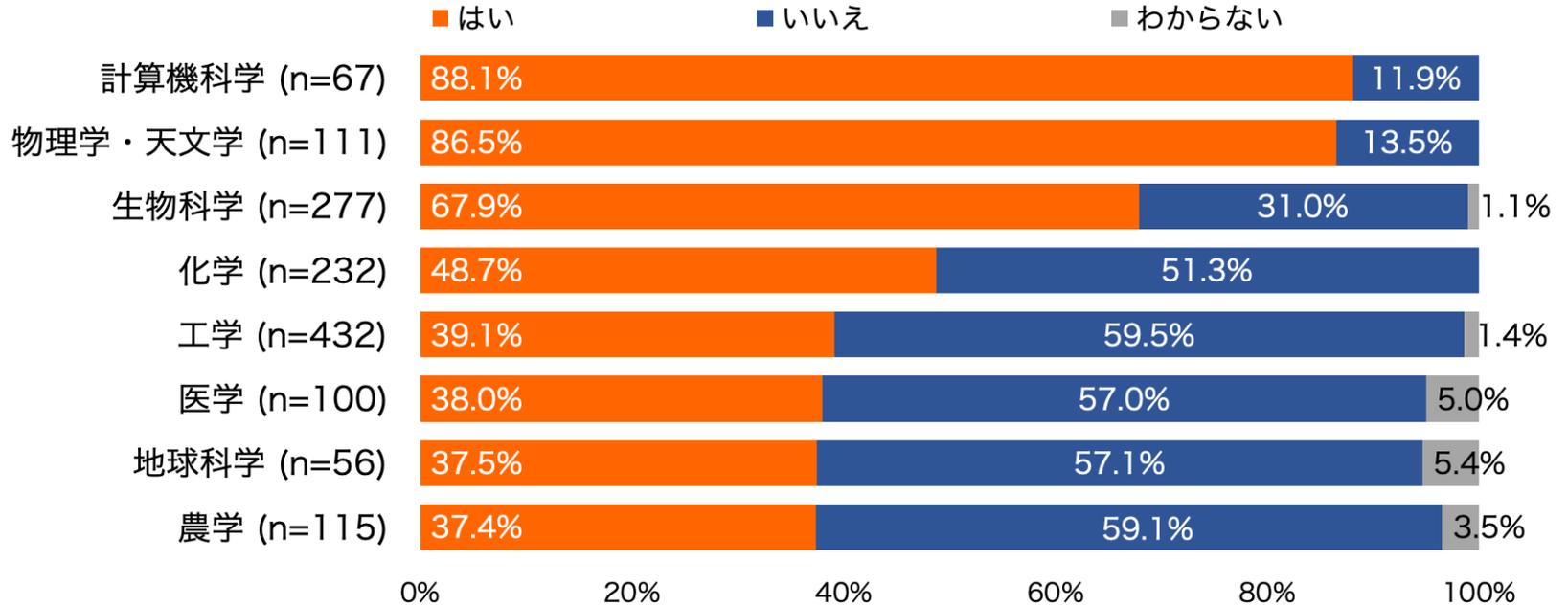


世代交代による変容を示唆

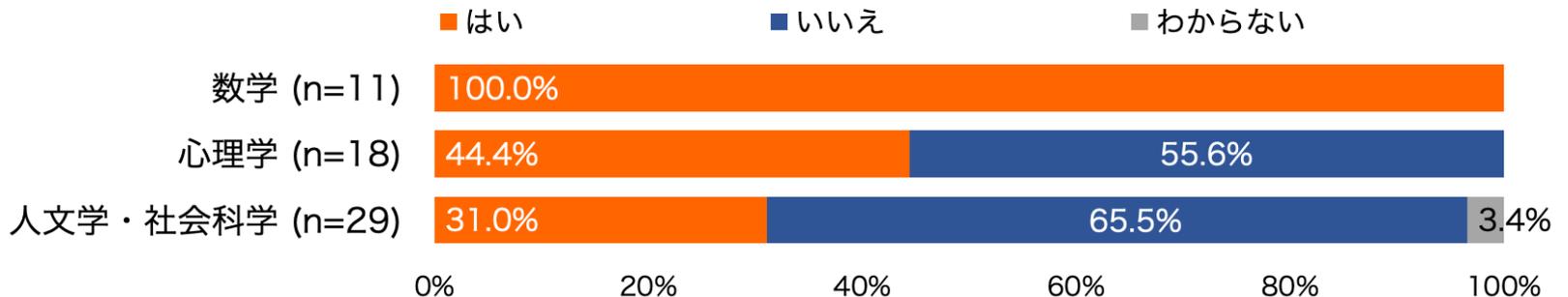
(n=1,447) ※不明(1名)を除く

[doi/10.15108/rm301](https://doi.org/10.15108/rm301)

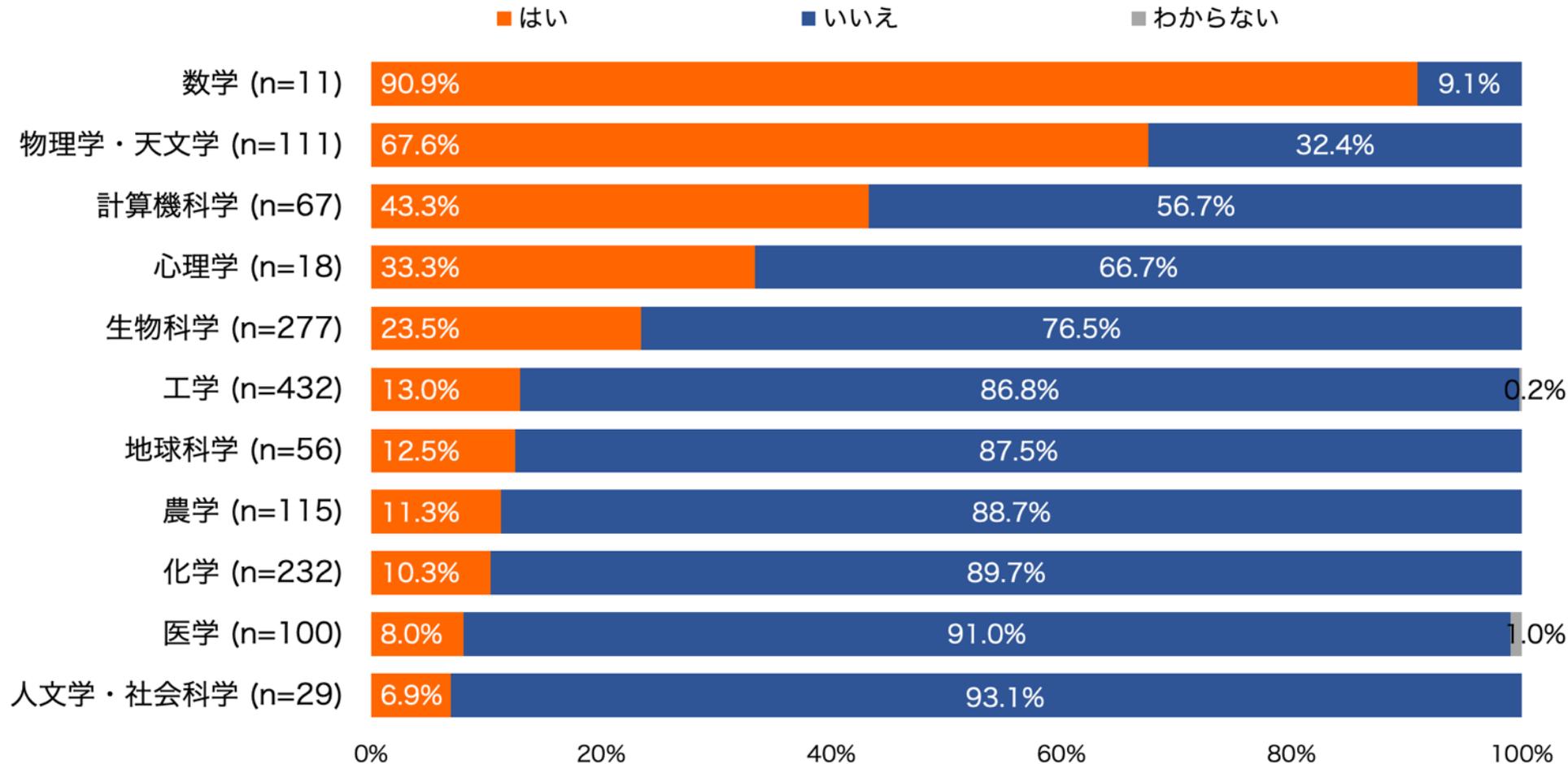
分野別プレプリントの入手経験（n=1,390）



参考：50名未満の分野（n=58）

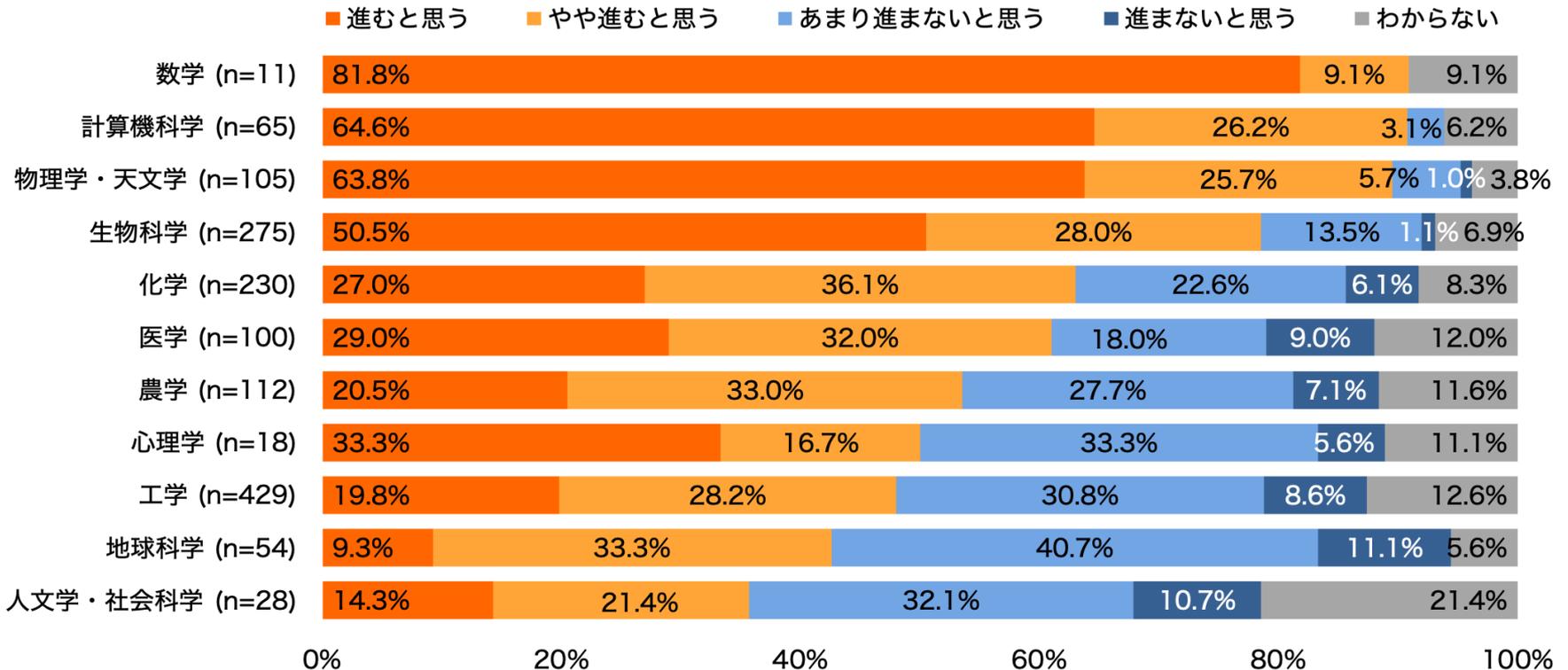


分野別プレプリントの公開経験（n=1,448）



分野別のプレプリントの展望

Q (自身に最も近い研究分野を選んだ後)
その分野では、今後プレプリントの利用が進むと思われますか。



(n=1,427)

多くの分野で進展を見込む割合が多い

■ 原著論文を書く研究者

- ◆ 自身の研究成果公開・共有の新しいツールとして無視はできない
- ◆ 戦略的活用
 - 先取権の確保, 共同研究者の募集, 集合知の活用
 - 公開のタイミングの是非

■ 原著論文を読む研究者, 専門家

- ◆ 問われる見識 (6割は査読に通るとして)
- ◆ なぜ高エネルギー物理学でプレプリントが浸透したか

■ その他の利用者

- ◆ 原則として取り扱いには慎重に (玉石混交, 撤回もあり)
- ◆ 安易に広報, シェアしない

■ リスクのとり方

■ プレプリントをジャーナルとしてどう扱うか

- ◆ 投稿規定等の修正
- ◆ 若手研究者へのアピールは必要か

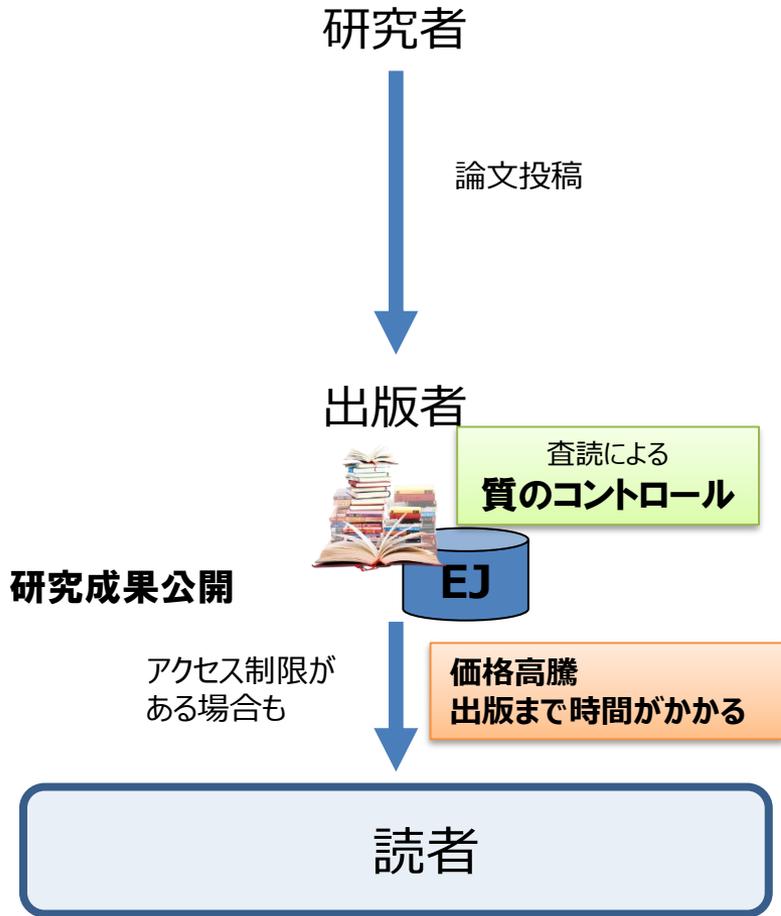
■ 査読付ジャーナルの編集・運用戦略への影響

- ◆ ダイレクトトランスファーへの対応（水引きは必要）
- ◆ プレプリントから良い論文を“青田買い”することになるか？
 - レビュー論文候補を早めに見つけるのは今からでもできる

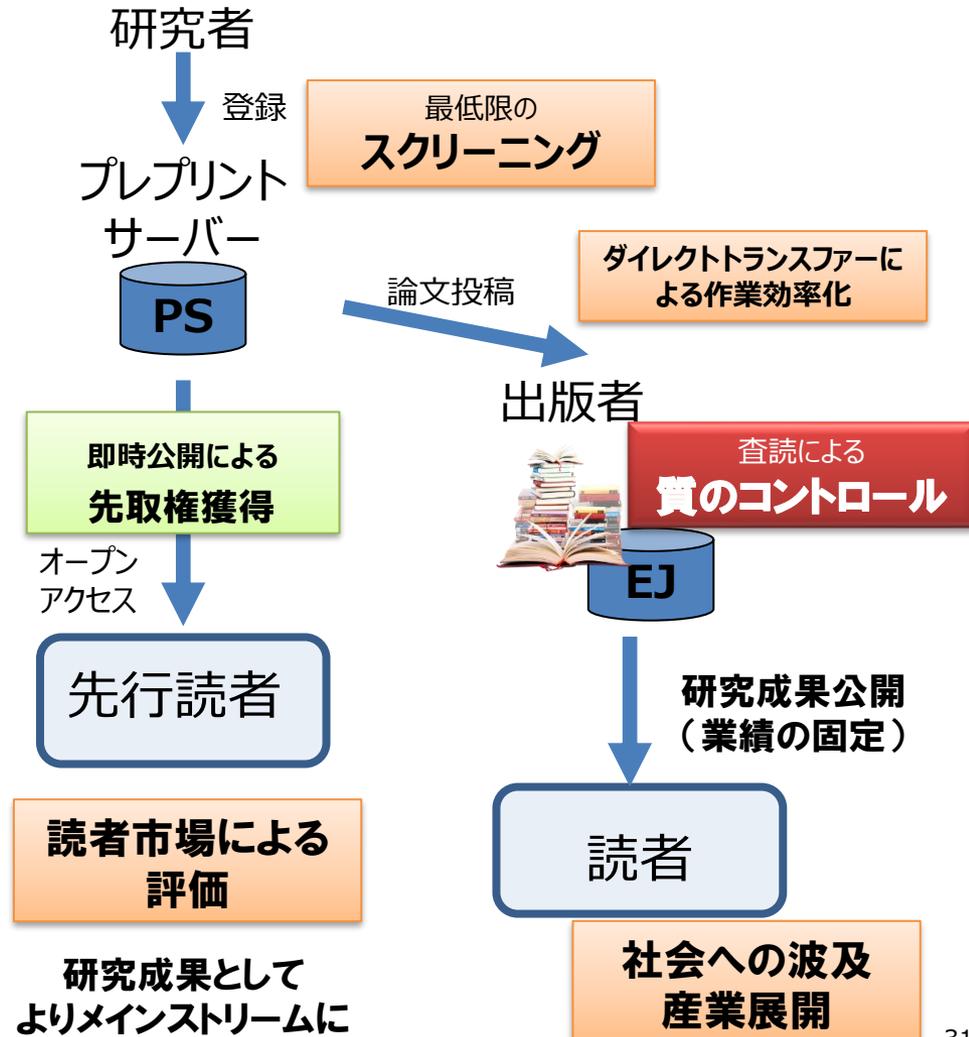
■ プレプリントの利用に関する啓発活動

- ◆ 特に社会向けにどこまで説明を尽くすか

従来の仕組み

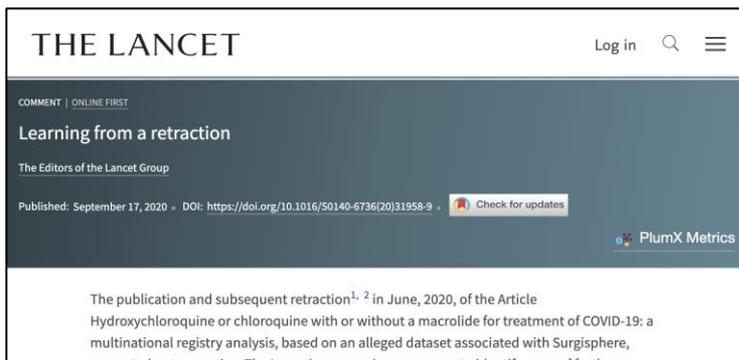


プレプリントサーバー



- **New England Journal of Medicine(NEJM)、Lancetという2大臨床医学誌でCOVID-19に関する論文が撤回**
 - ◆ NEJMの方は、新型コロナと心血管疾患や薬剤との関係を検討した臨床研究
 - ◆ Lancetの方はクロロキン、ヒドロキシクロロキンという新型コロナに対する治療薬に関する観察研究
- **これら2つの論文はいずれの論文もサージスフィア (Surgisphere) という社員わずか数名のデータ分析会社のデータによるもの。**
- **両論文の疑義を明らかにするためにデータの開示を求められたところ、サージスフィア社はデータ開示を拒否**
- **再検証は困難であるということで論文撤回**

<https://news.yahoo.co.jp/byline/kutsunasatoshi/20200606-00182086/>

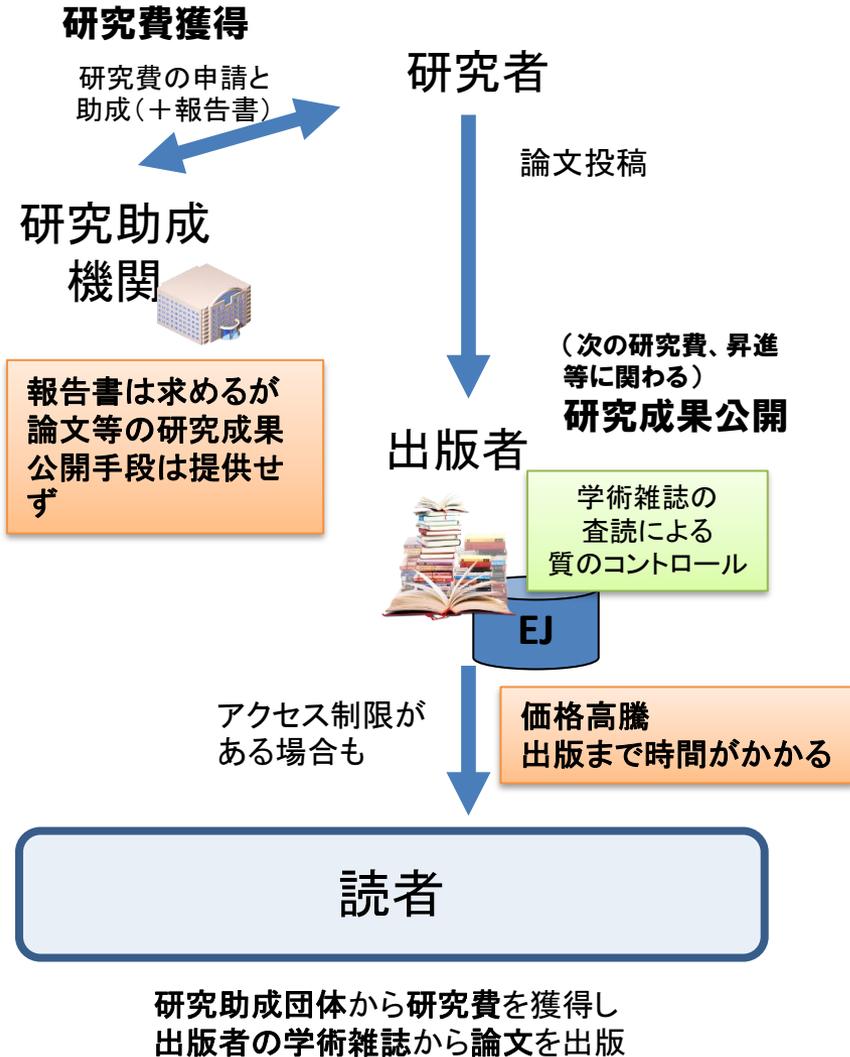


COVID-19の論文撤回を踏まえて

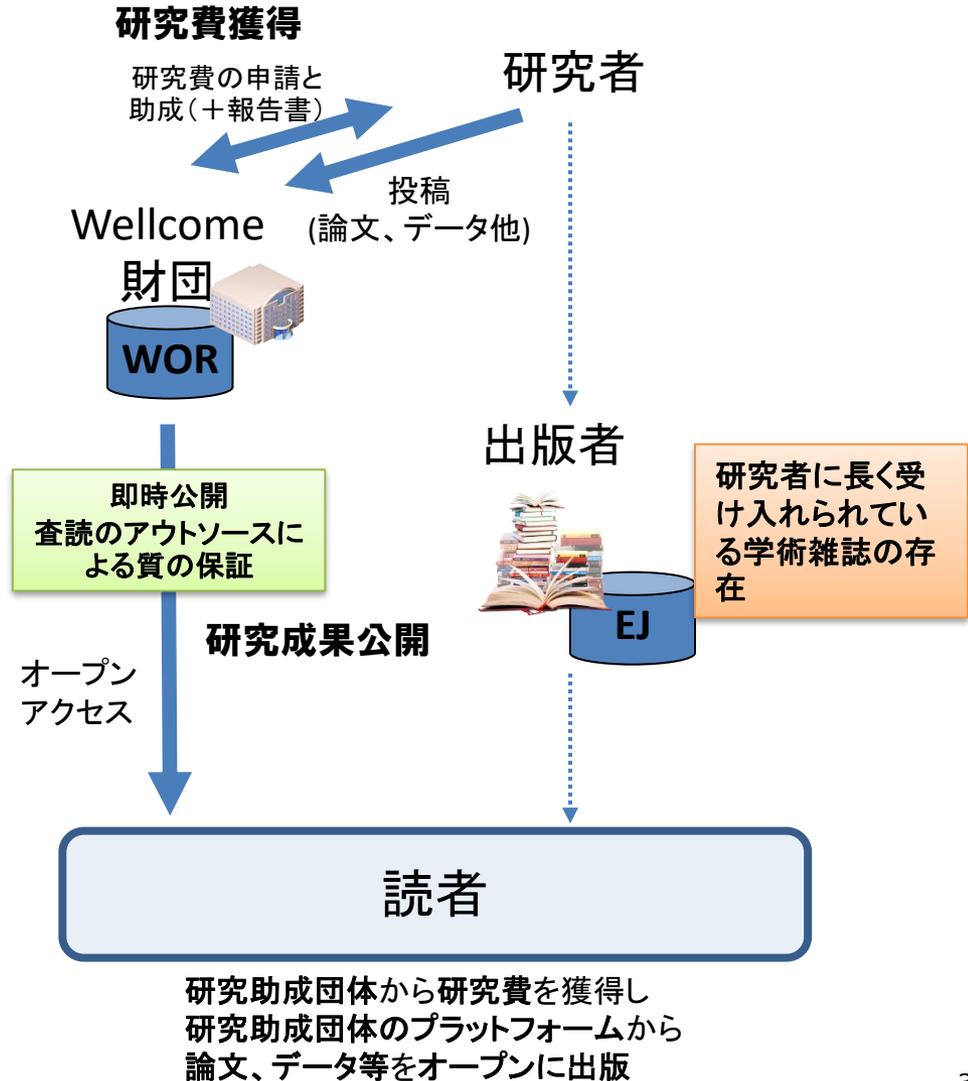
- 論文の査読者の専門知識の基準を厳格化 (データサイエンティストの導入)
- すべての著者がデータの妥当性を保証
- データ共有計画を詳細に記述させる

- [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(20\)31958-9/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)31958-9/fulltext)

従来の仕組み



Wellcome財団の試み (Wellcome Open Research: WOR)



5. まとめ

- プレプリントの共有は昔から行われており、インターネットとwebによって、世界中に草稿を迅速に共有できる時代になった
- プレプリントサーバー自体には30年以上の歴史があるが、COVID-19によって医学系他の領域においても一気に注目を集め、世界的な社会課題解決のためなど、プレプリントならではの価値の提供ができるようになった。
- プレプリントの利活用の是非は立場によって変わり、大まかには、研究者（情報発信）としては積極的かつ戦略的に利用し、読者（受信）としては、リスクを理解した上で慎重に取り扱うべきものである。
- プレプリントの進展は、学術雑誌と原著論文を中心とした研究者のエコシステムのあり方に一石を投じており、オープンサイエンスが指向する科学と社会の変容を駆動するものである。