



# プレプリントが変える学術情報流通と オープンサイエンスの展望

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

AMED科学技術調査員

日本学術会議特任連携会員

2023年9月28日(金)

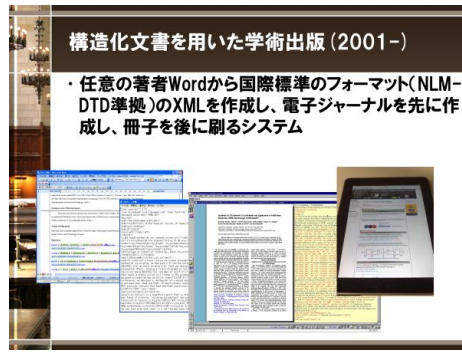
UniBioPress シリーズ学術出版を学ぶ(28)－学術情報流通の未来



## CONTRIBUTION

- 1990
  - Organic Syntheses
- 1995
  - Digital Peer-Review Tracking System
- 2000
  - Online journals
  - XML Publishing
  - DOI and CrossRef
  - EJ-business
- 2005
  - Open Access 2005
  - Advocacy for policy makers
- 2010
  - ORCID
  - altmetrics
- 2015
  - RDA Tokyo 2016
  - G7 OSWG
  - OECD
  - Citizen Science
- 2020
  - Blockchain
  - Virtual Learned Society
  - DeSci

科学研究をしながら、紙からデジタル、デジタルからweb(インターネット)へのメディア変容について実際に手を動かす



SGM, XML scholarly publishing

Advocacy for Science Council of Japan



DX for PTA



Citizen Science (NHK)

オープン化の潮流、論文からデータへの拡張



International Contribution For Open Science Policy



Japan Open Science Summit



Scholarly communication on Metaverse



Blockchain taskforce for Chemistry



DX for Learned Society

科学のボーダーレス化(シチズンサイエンス) DeSciやバーチャル学会の立ち上げをサポート

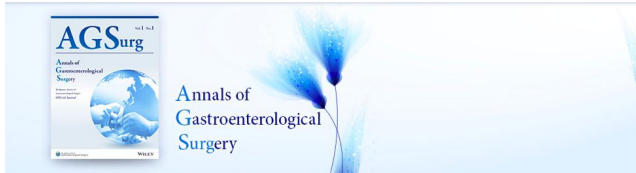


Decentralized Science blockchain

# From Bottom-up to Top-down



Chemical Society of Japan (1995-2012)  
 Journal Manager  
 EJ development (with my IT Skill)  
 OA implementation  
 ALPSP Board Member (2011)



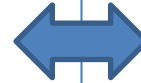
<https://onlinelibrary.wiley.com/journal/24750328>



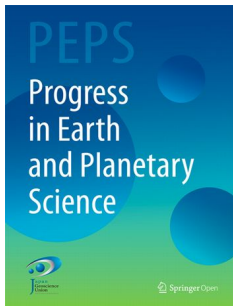
<https://www.cell.com/patterns/>



National Institute of Science and Technology Policy (2012-)  
 Open Science policy development



GAP analysis  
 Translation  
 Consultation



<http://progearthplanetsci.org/>

Advisory Board Member  
 Consultation



<https://iupac.org/>



Expert Member, Advisory Committee

実際に電子ジャーナル開発と運営ならびにOA化を経験した研究者が、オープンサイエンス、研究データ共有の政策づくりに携わり、変容を駆動する

- SPARC Japan、J-STAGE
- XSPA(学術XML推進協議会)
- 科研費成果公開促進費改定



- Japan Open Science Summit
- RDUF(研究データ利活用推進協議会)
- AMED情報分析課

- プレプリントは査読付き原著論文の草稿であるが、研究成果公開メディアとして一定の独立した地位を他分野で確立しつつある。
- **プレプリントの利活用の是非は利用者の立場によって変わる。**
- 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いる。
- **プレプリントがもたらす研究および研究環境の変化に応じた対応が求められる。**

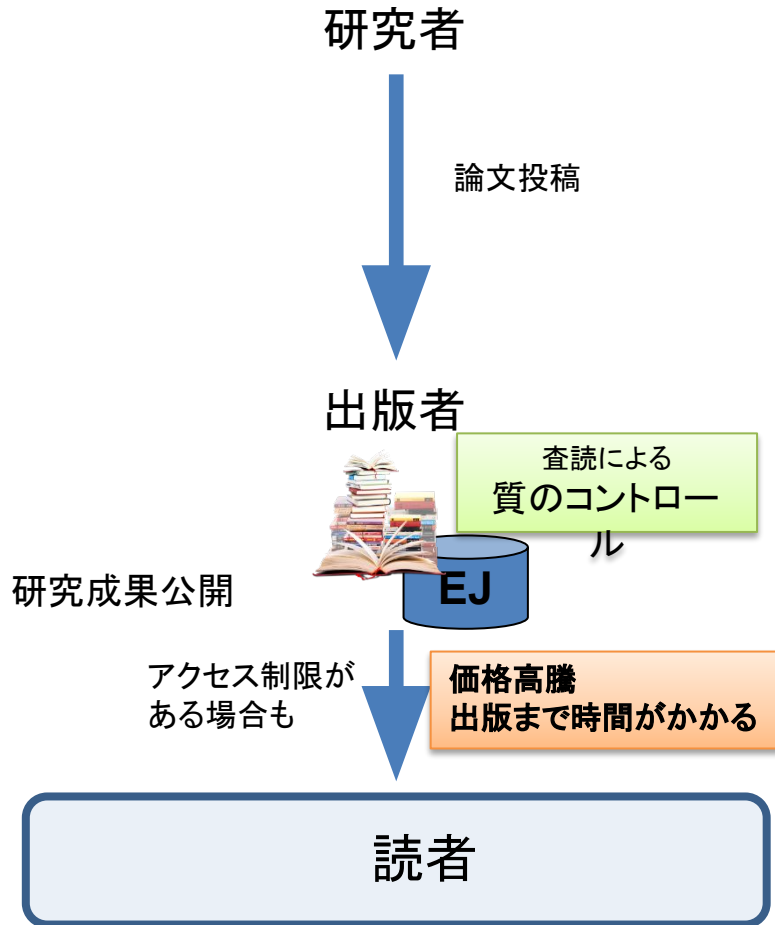
- **プレプリント＝査読つきジャーナルに通る前の草稿原稿**
  - ◆ プレプリントは査読済みでもなく、出版(Publish)されたものでもない
  - ◆ 紙の時代からあるプレプリントの共有(身内, 共同研究者)
  
- **プレプリントサーバー＝プレプリントを掲載するインターネットサーバー(Web)**
  - ◆ 1991年から30年あまりの歴史(物理)
  - ◆ “身内と共有”から“世界と共有”へ



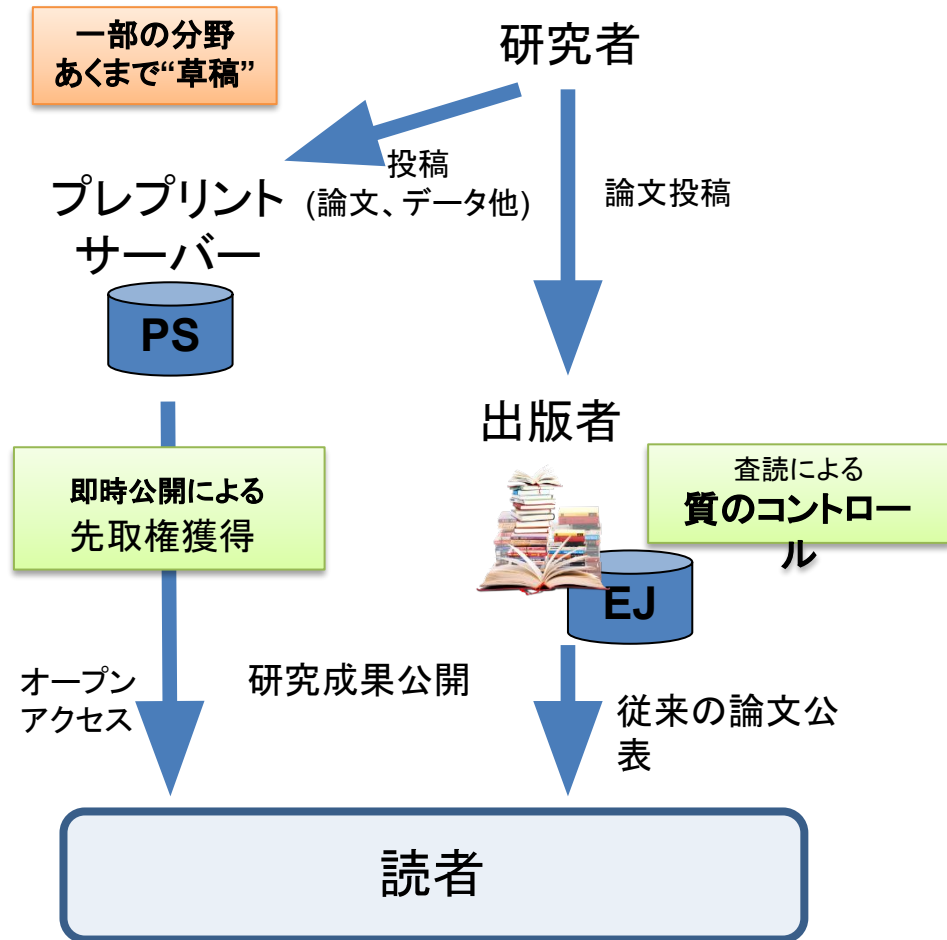
<https://www.medrxiv.org/>

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

## 従来の仕組み



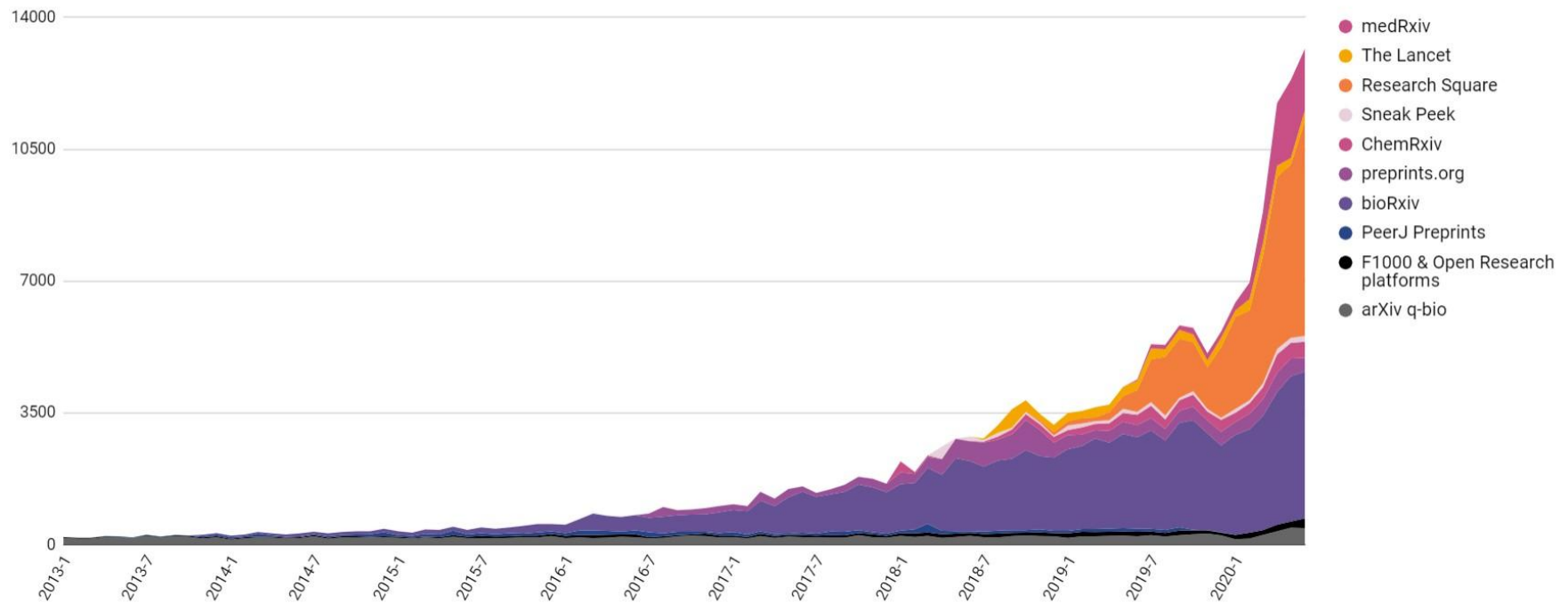
## プレプリントサーバ(PS)の活用



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

Biomedical preprints per month through 2020-06

Sources: Jordan Anaya (PrePubMed), Naomi Penfold, EuropePMC, arXiv, Crossref, SSRN



Polka, J K., & Penfold, N C. Biomedical preprints per month, by source and as a fraction of total literature (4.0). Zenodo. [10.5281/zenodo.3955154](https://doi.org/10.5281/zenodo.3955154)

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

## ■ 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/>

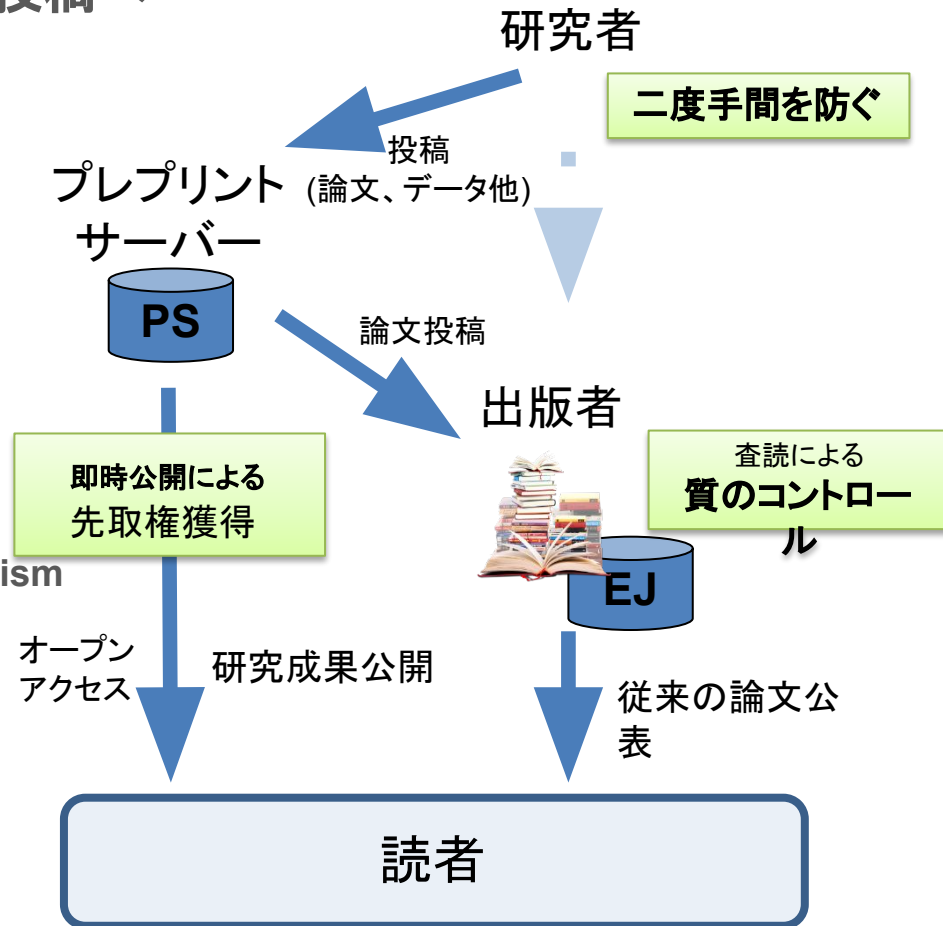
2019年6月5日発足

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.**

## プレプリント登録からそのままジャーナル投稿へ

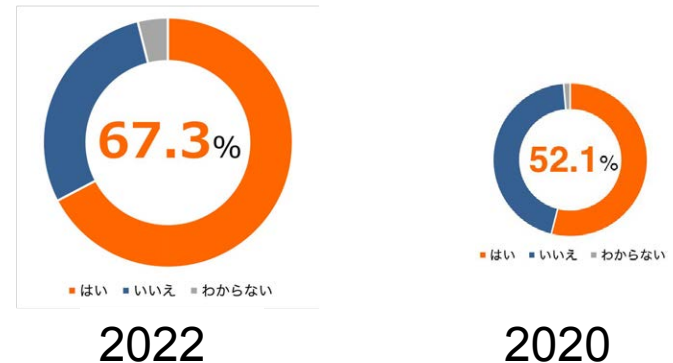
- BMJ Open
- Cells Tissues Organs
- Developmental Neuroscience
- EMBO Molecular Medicine
- Endocrinology
- Genetics in Medicine
- Human Heredity
- Intervirology
- JCO Clinical Cancer Informatics
- JCO Precision Oncology
- The Journal of Clinical Endocrinology & Metabolism
- Journal of the Endocrine Society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



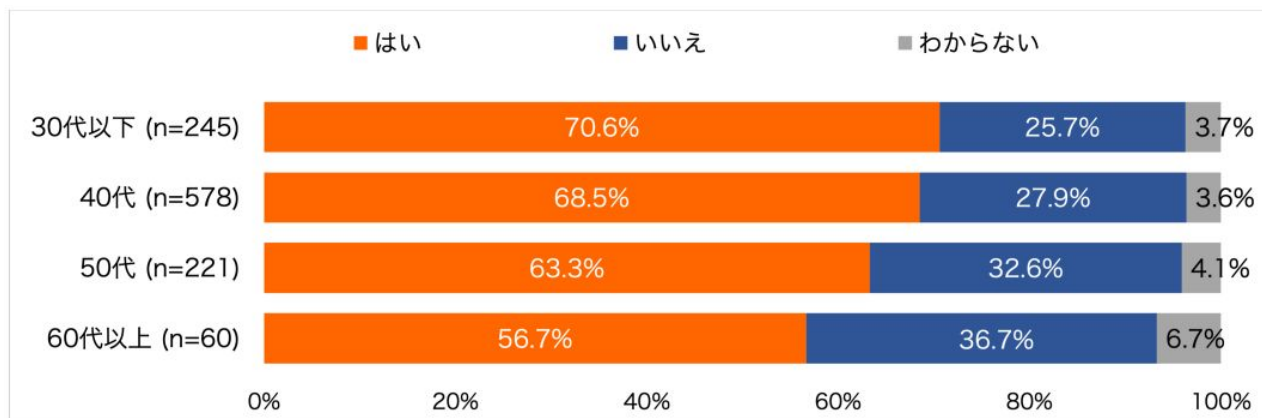
- 投稿者: 1回の投稿作業で、プレプリントと査読付きジャーナルへの作業が完了
- 出版社: ある程度のスクリーニングが済んだ原稿を査読できる

- 調査対象
  - 科学技術専門家ネットワーク:1,671名
  - 有効回答:1,173名(回答率70.2%)
- 調査期間
  - 2022年7月19日～8月22日

## プレプリントの入手経験(全体)



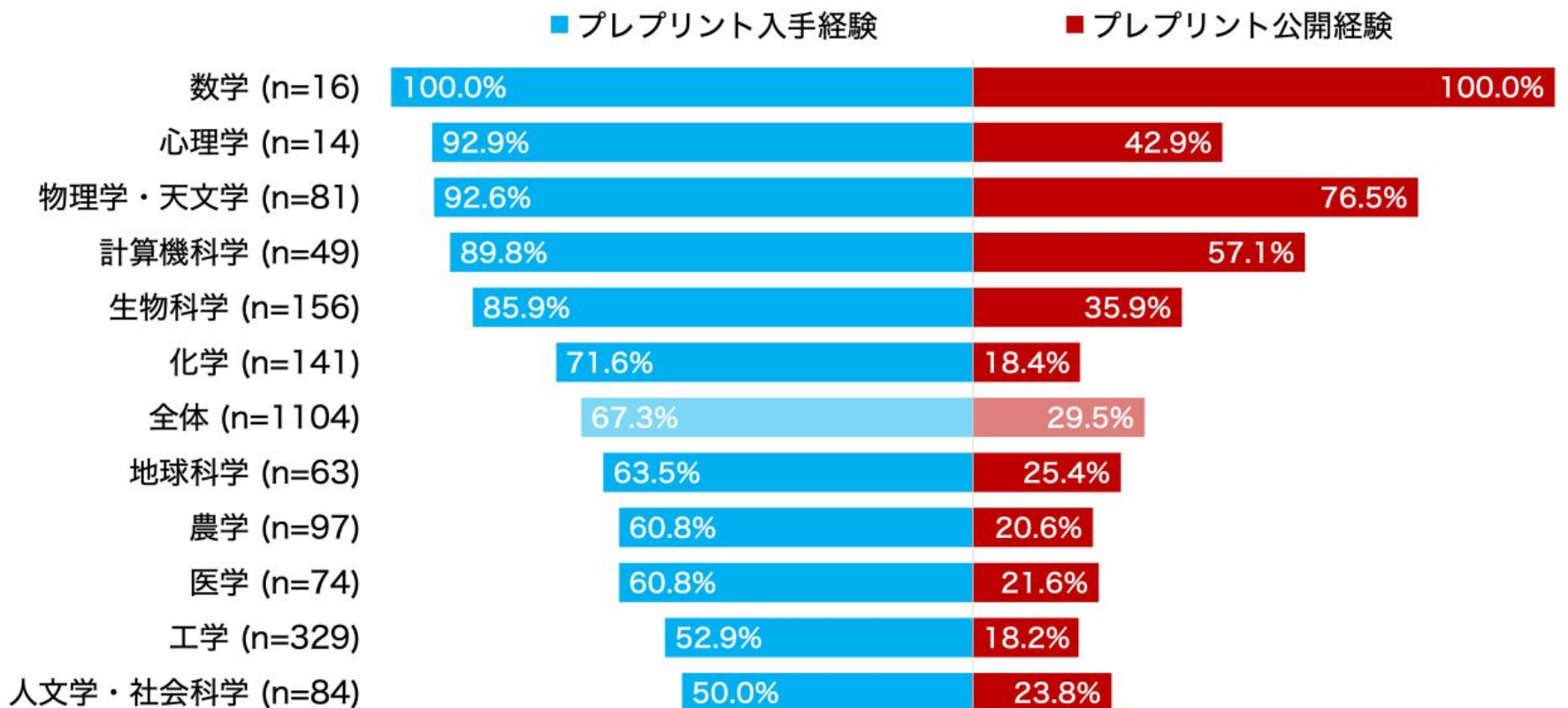
## 年齢層別プレプリントの入手経験



世代交代による変容を示唆

(n=1,104)

## 分野別プレプリントの入手経験 (n=1,104)

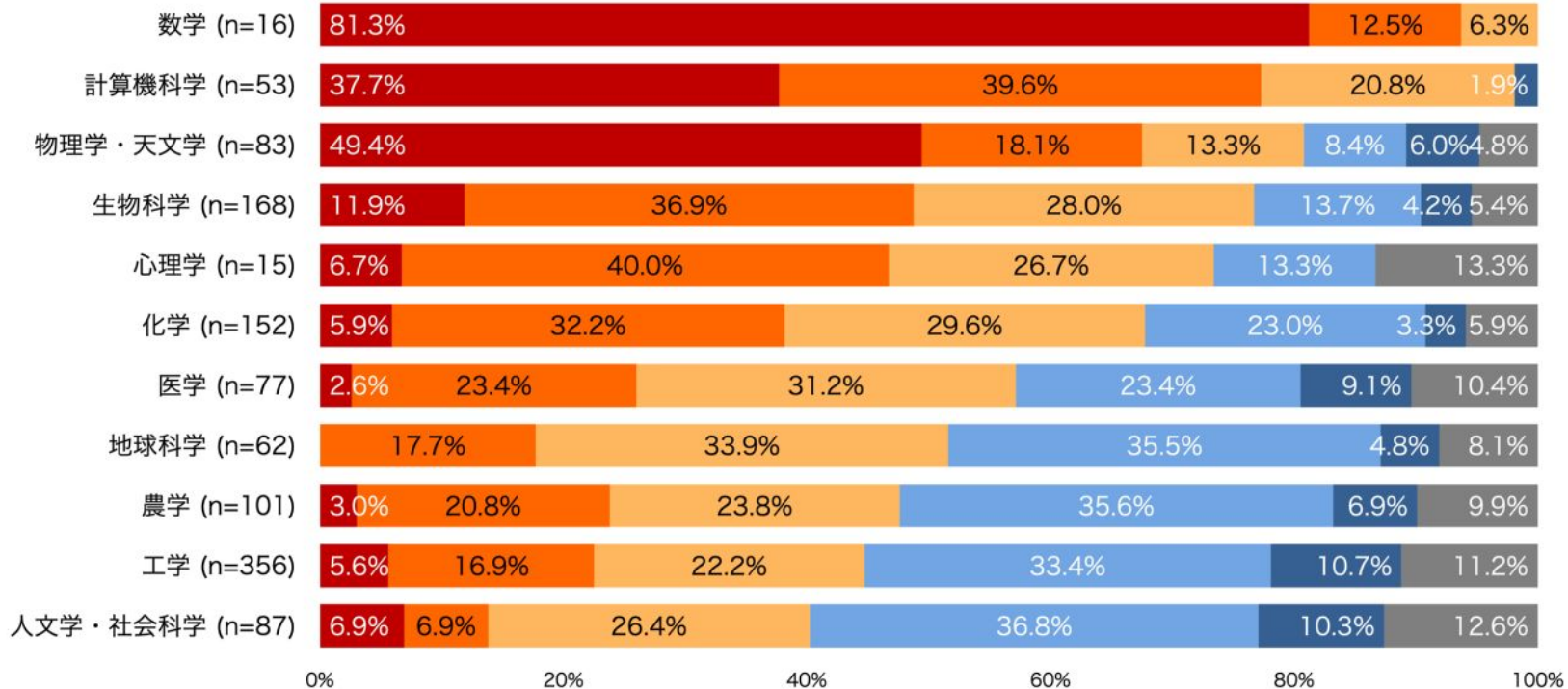


入手: 38.0%(2020)から大幅増  
 公開: 8.0%(2020)から大幅増

# 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。

- 既に十分利用されていると思う
- 進むと思う
- やや進むと思う
- あまり進まないと思う
- 進まないと思う
- わからない



医学を含む多くの分野で進展を見込む割合が多い (n=1,104)

## ■ Global Challenge (世界的な社会課題)

- ◆ COVID-19
- ◆ ジカ熱
- ◆ SARS

緊急対応としての共有

## ■ Super Hot Topic (注目の高いトピック)

- ◆ 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)

## ■ Super Novel Topic (新奇性が高いトピック)

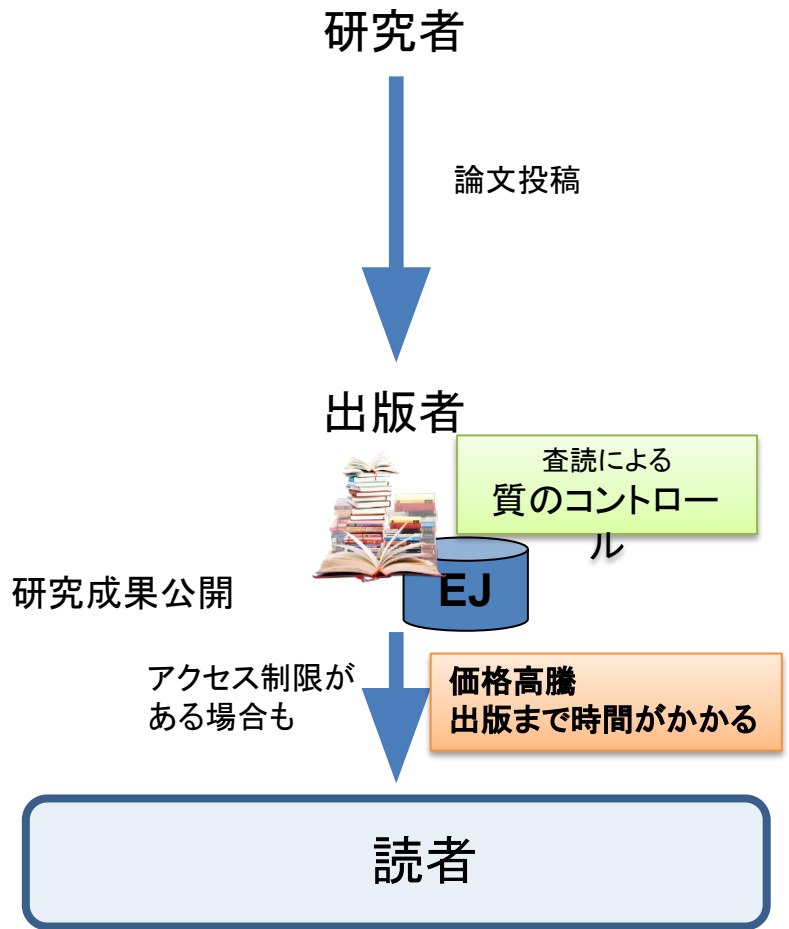
- ◆ 先進的過ぎると査読に通らない(時間がかかりすぎる)、査読者が居ない ポアンカレ予想, ABC予想

## ■ Super Societal Impact (実用性の高いトピック)

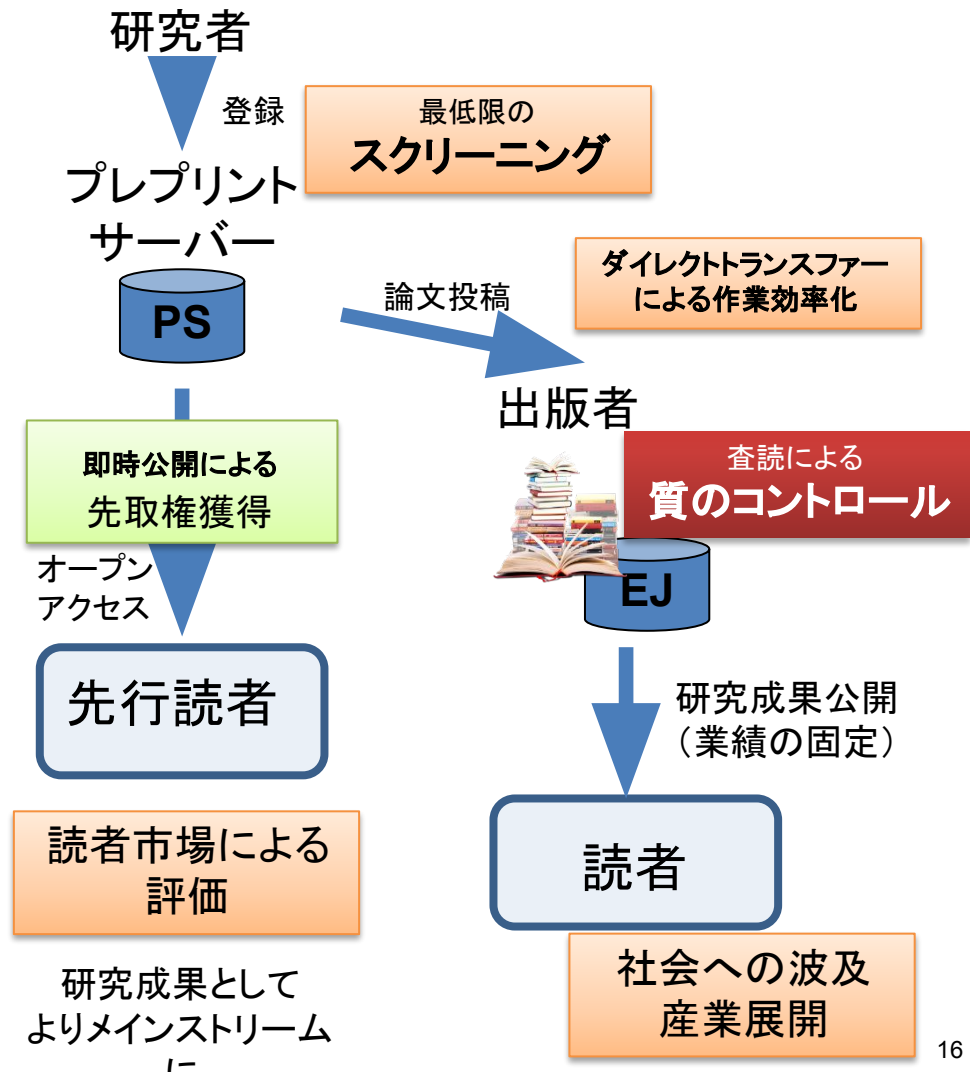
- ◆ 実用性が高いものを手早く共有 Google Page Rank

- **原著論文を書く研究者**
  - ◆ 自身の専門における研究成果公開・共有および最新の情報収集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(プレプリントを出さない戦略もあり)
- **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は慎重に(玉石混交, 撤回もあり)
  - ◆ 安易に広報(含むプレスリリース), シェアしない
- ✓ **リスクのとり方**

### 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/>

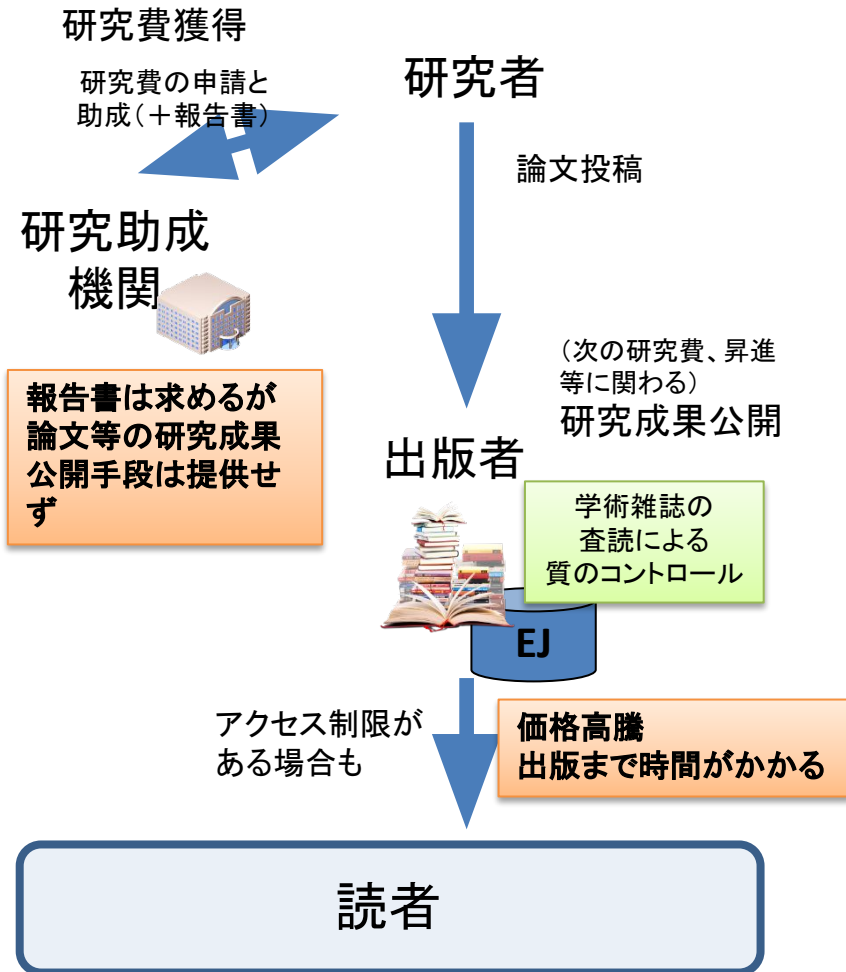


COVID-19の論文撤回を踏まえて

- 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

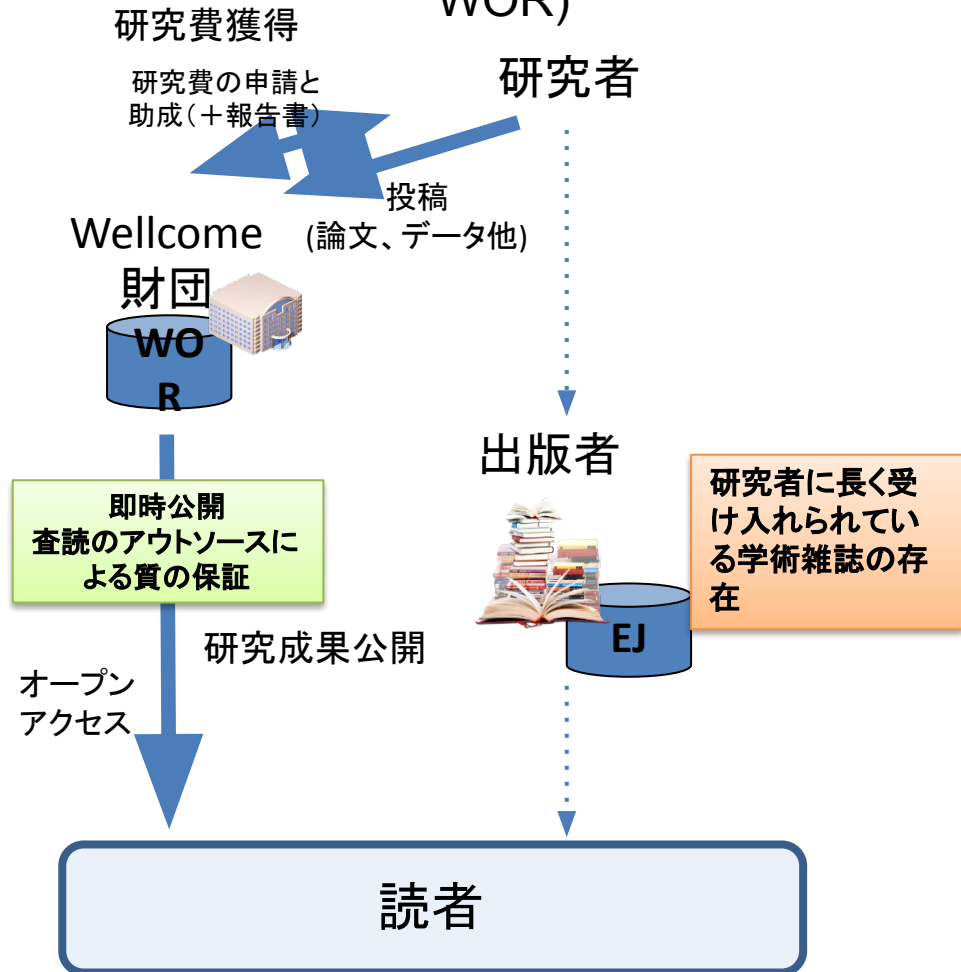
- [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(20\)31958-9/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)31958-9/fulltext)

## 従来の仕組み



研究助成団体から研究費を獲得し  
出版者の学術雑誌から論文を出版

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



研究助成団体から研究費を獲得し  
研究助成団体のプラットフォームから  
論文、データ等をオープンに出版

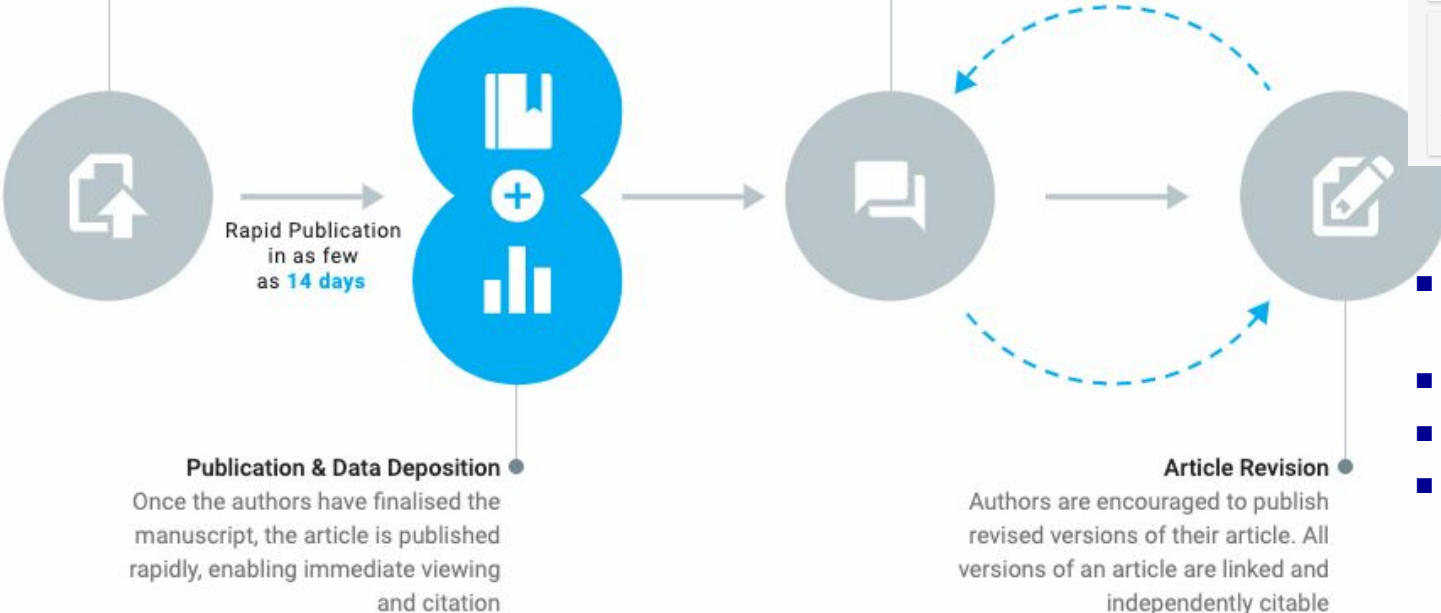
## Open Research Central The Publishing Model

### Article Submission

Submission is via a single-page submission system. The in-house editorial team carries out a comprehensive set of prepublication checks to ensure that all policies and ethical guidelines are adhered to.

### Open Peer Review & User Commenting

Expert peer reviewers are selected and invited, and their reports and names are published alongside the article, together with the authors' responses and comments from registered users



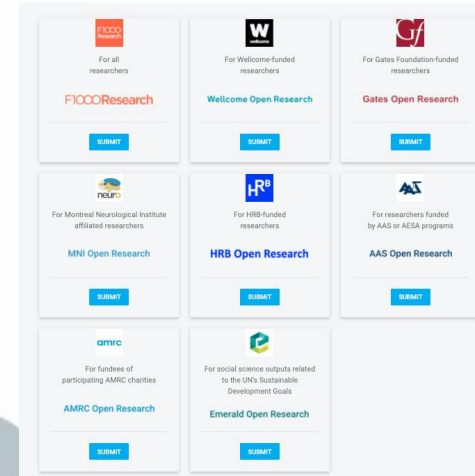
Rapid Publication  
in as few  
as 14 days

### Publication & Data Deposition

Once the authors have finalised the manuscript, the article is published rapidly, enabling immediate viewing and citation

### Article Revision

Authors are encouraged to publish revised versions of their article. All versions of an article are linked and independently citable



- F1000Research (筑波大学)
- ウェルカムトラスト
- ゲイツ財団
- EUで導入検討の動きも

- 論文を書いたら、データと共にまずは公開(出版)
- 後から透明性の高い査読と、ユーザーコメント
- 改訂を加えてブラッシュアップ

## ■ eLife



プレプリントとの境界があいまいに

<https://elifesciences.org/inside-elifesciences/54d63486/elifesciences-s-new-model-changing-the-way-you-share-your-research>

- ◆ 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)

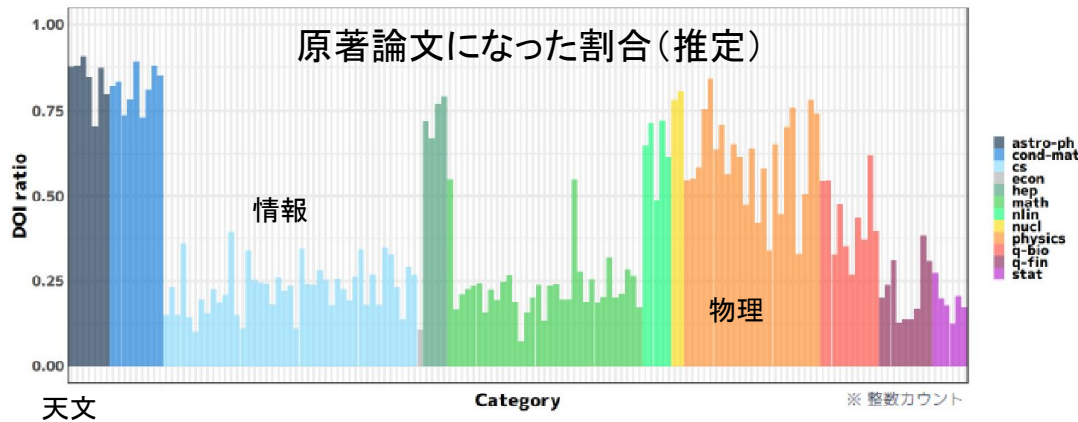


図5 分野毎のDOI付与率

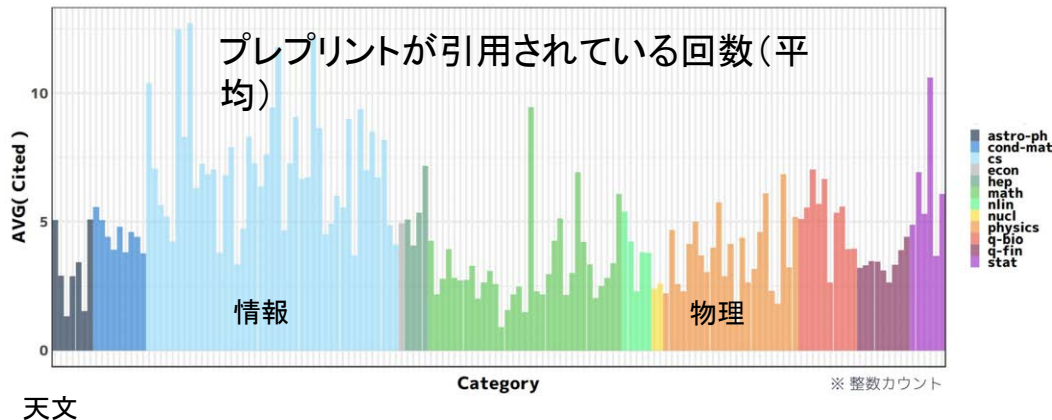
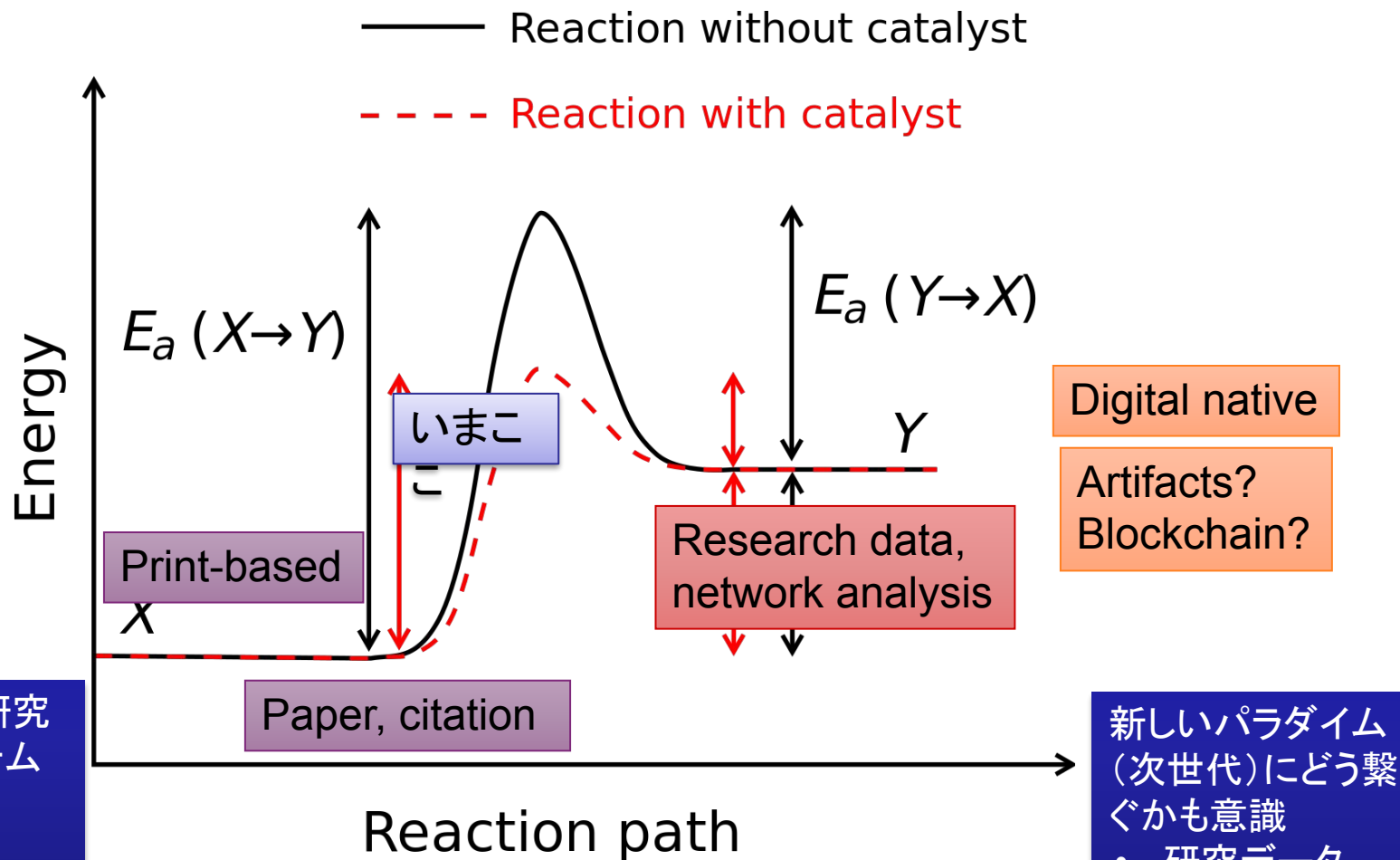


図7 分野と被引用の関係性

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

# 遷移状態をどう切り抜けるか (楽しめるか)



Base chart: [https://commons.wikimedia.org/wiki/File:Activation\\_energy.svg](https://commons.wikimedia.org/wiki/File:Activation_energy.svg)

これまでの研究のエコシステムは依然重要

- OA
- プレプリント
- 転換契約

新しいパラダイム (次世代) にどう繋ぐかも意識

- 研究データ
- 定量評価
- 研究DX

# 新たなオープン化(知の開放)に基づく社会制度と、方針と運用の再デザイン(新しい秩序)

大量印刷と物流が  
支えてきた科学と社会

Webが支える  
科学と社会

Human Readable

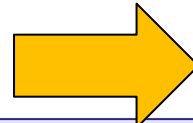


Past Design

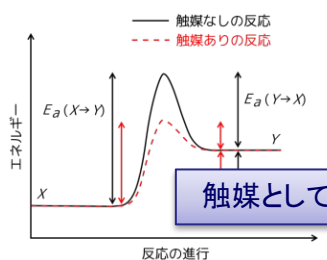
Open  
Close  
Secret

Chubin(1985)

過去から引き続き  
社会制度に応じた  
対応方針、運用



情報爆発  
による知の開放



[http://commons.wikimedia.org/wiki/File:Activation\\_energy\\_ja.svg](http://commons.wikimedia.org/wiki/File:Activation_energy_ja.svg)



Machine Readable

Future Design

Open  
Close  
Secret

これからの  
社会制度に応じた  
対応方針、運用

EC,  
OECD

ICTは進展したが、著作権や知財を含む法律、  
社会制度の骨格は旧来のまま

新オープン・クローズ戦略

- 科学・知財を取り巻く(人の行動原理を中心とした)本質は同じだが、情報基盤の変革に応じた再デザインと新しい秩序形成

- プレプリントは査読付き原著論文の草稿であるが、研究成果公開メディアとして一定の独立した地位を(医学系でも)確立しつつある。
  - ◆ 特にホットピックの研究者間の共有に関しては、事実上不可欠
  - ◆ 原著論文とは違うメディアとしての役割の可能性が出てきた
- **プレプリントの利活用の是非は利用者の立場によって変わる。**
  - ◆ 研究者の情報収集と発信においては自身の専門性のもと積極的かつ戦略的にプレプリントを活用
  - ◆ 非専門の読者としてはリスクを踏まえて慎重に取り扱うべきものである
  - ◆ 研究者コミュニティ(学会)としての見識、啓発が求められる
- **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、オープンサイエンスパラダイムに向けた変容を促している。**
  - ◆ (短期的)査読、雑誌のあり方
  - ◆ (長期的)研究、知識形成のあり方

(以降参考資料)

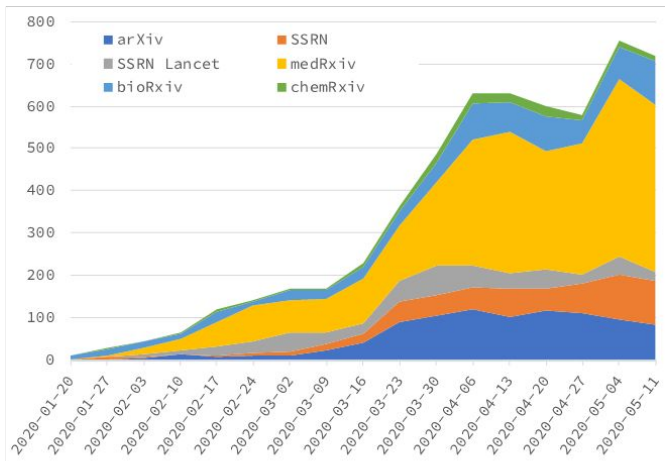
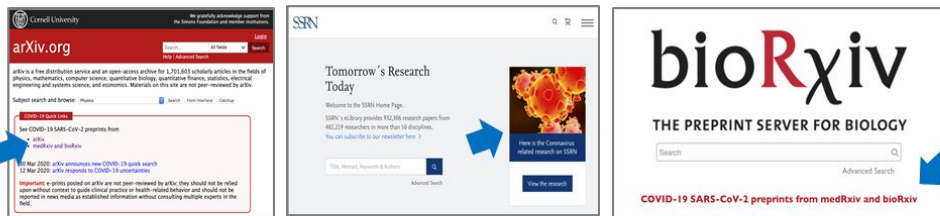
- **1969年インゲルフィンガー・ルール (Ingelfinger Rule)**
  - ◆ その内容が既に投稿されていたり、どこか別の場所で報告されていたりした場合は、原稿の受け取りを拒否する
- **1999年E-Biomed by H. Varmusが理論物理学のプレプリントサーバーを真似る形で医学版arXivを提唱するも頓挫**
  - ◆ 内容を理解できる人が限られる理論物理学とは異なり、健康や生命に直接的にかかわり多くの人が興味を持つ医学医療分野の懸念
  - ◆ これを機にPMC (査読済み論文のポータル) 創立、PLoS ジャーナル (Gold OAジャーナル) 創刊へ
-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

- 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



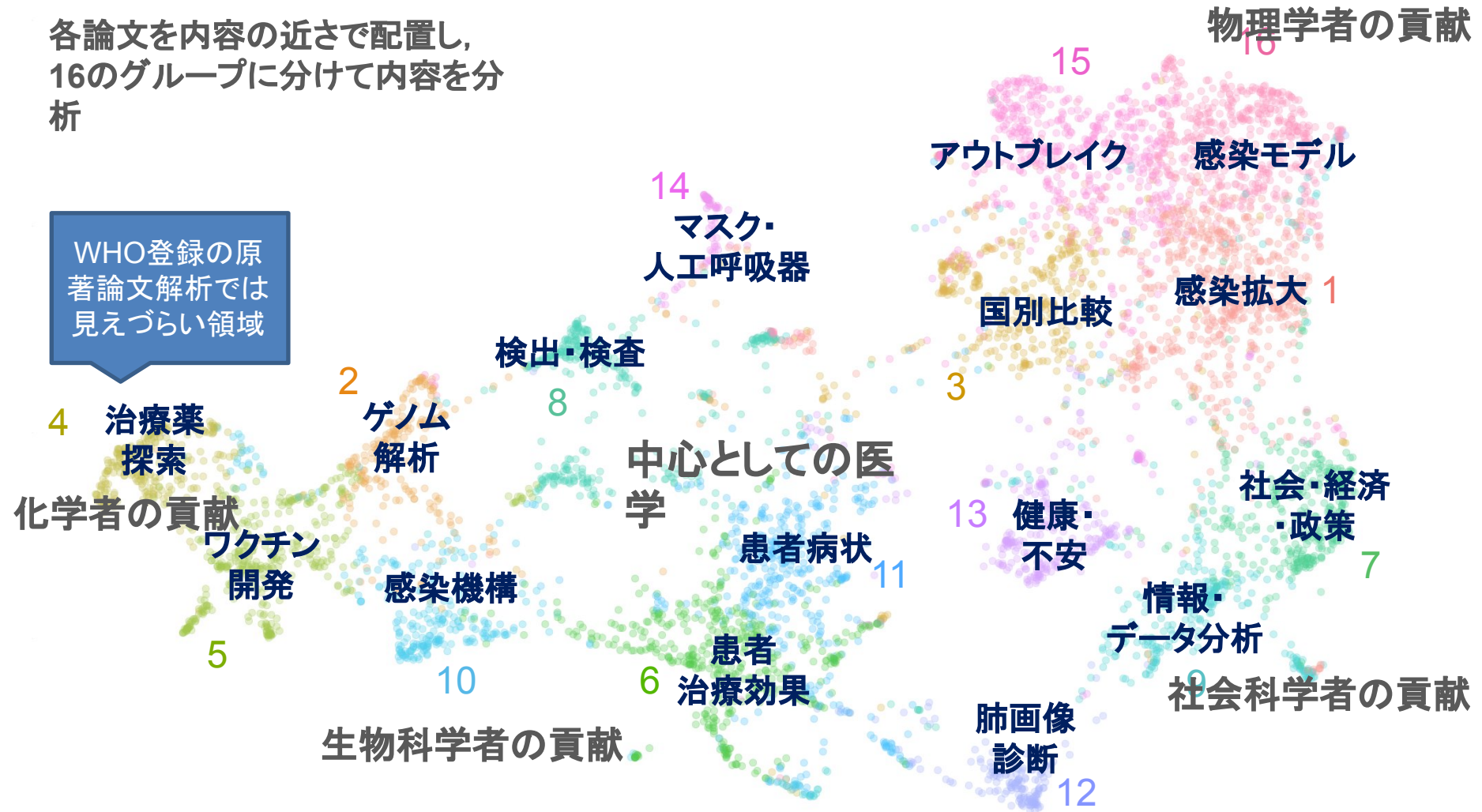
Week	arXiv	SSRN Lancet	medRxiv	bioRxiv	chemRxiv
2020-01-20 (04)	0	1	0	9	1
2020-01-27 (05)	2	5	1	16	1
2020-02-03 (06)	4	5	5	16	1
2020-02-10 (07)	13	2	8	27	5
2020-02-17 (08)	7	4	20	59	6
2020-02-24 (09)	10	6	28	84	9
2020-03-02 (10)	10	10	45	76	5
2020-03-09 (11)	24	13	27	81	22
2020-03-16 (12)	42	20	26	106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

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

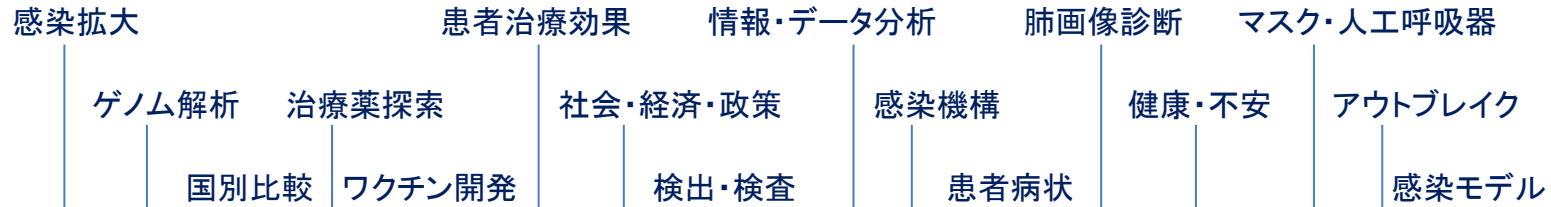
2020.05.18にデータを収集し, 2020年第20週 05/17 までのデータを取得  
(各PPSともに記事のPosted Dateを基準として採用)

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

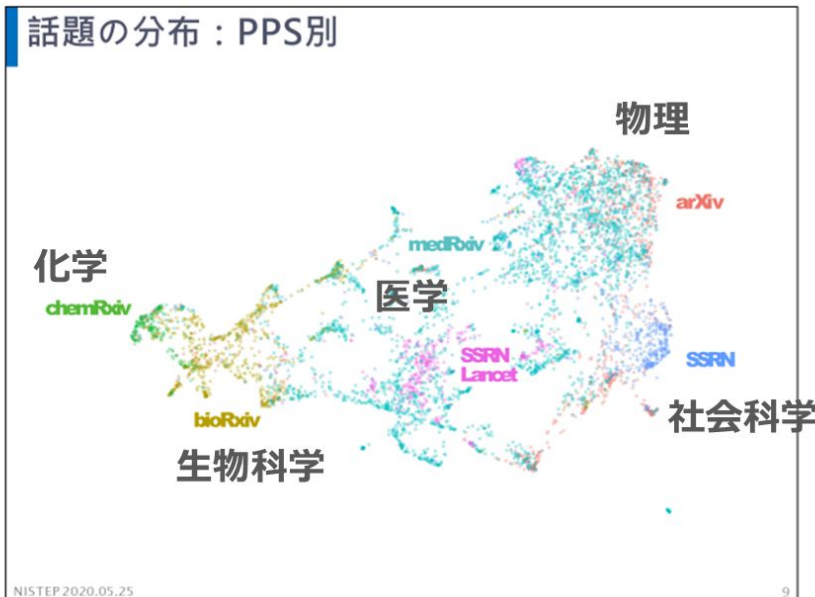
WHO登録の原著論文解析では見えずらい領域



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)