



COVID-19が加速する 学術ジャーナルと論文の変容

2020年11月13日

文部科学省科学技術・学術政策研究所

上席研究官 林 和弘

第4回研究大学コンソーシアムシンポジウム

パラレルセッション3 「新しい研究情報発信の在り方と課題」

経歴（研究情報発信の観点から）

1995 査読システム開発（+ SGMLの運用）

2000 独自システムからJ-STAGEへの転換

2001 メタデータファーストの学術出版確立

2002 DOI付与 Chemical Abstractとの連携

2005 OA対応 Googleとの連携

2009 dimensionsやe-CSTIの予察（ID付与とネットワーク分析の時代）

2010 XML出版確立と電子書籍対応

2011 ORCID啓発

2013 Altmetrics啓発

2015 オープンサイエンス啓発

2016 研究データ利活用協議会立ち上げ

2018 Japan Open Science Summit立ち上げ



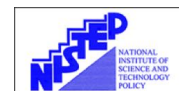
東大理化

[日本化学会
(アルバイト)]



日本化学会

[NISTEP（客員）]



NISTEP（常勤）

1990年代より、学術情報流通のDXに取組み、論文のを越えた、研究アイテムの拡張、IDの付与、及び、そのネットワーク化を啓発

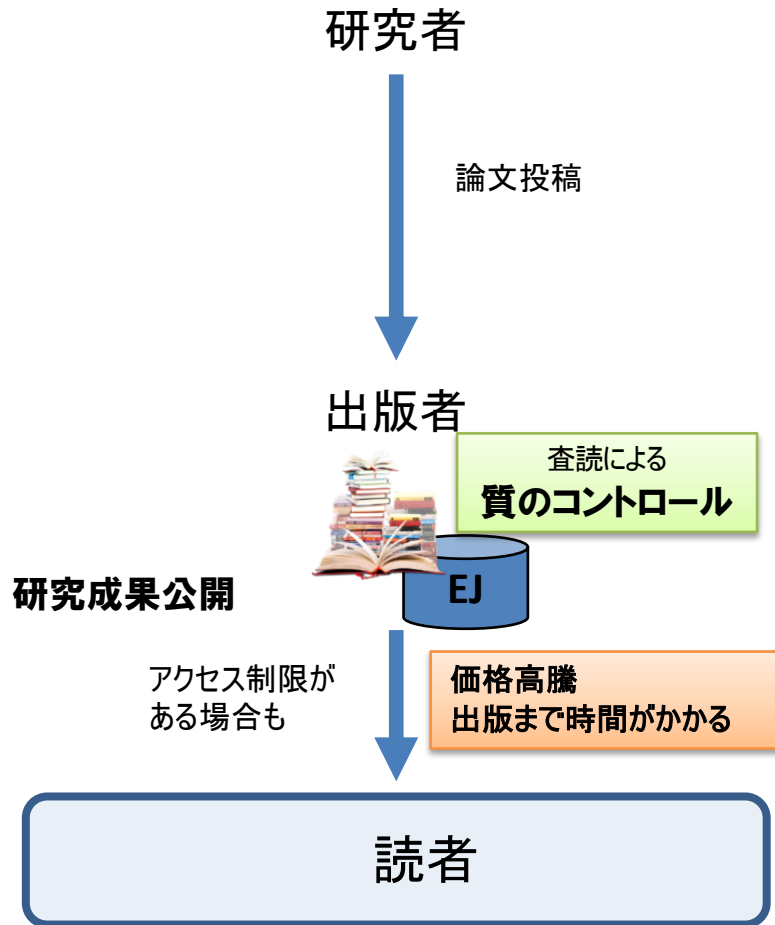
■ “学術ジャーナルと論文”のデジタルトランスフォーメーション

- ◆ 成果公開メディアの変容：プレプリントの可能性
- ◆ アクターの変容：新たなアクターと，出版者等が直接関与しないフロー
- ◆ “出版”という意味の再考：研究プロセスの各段階の共有と出版

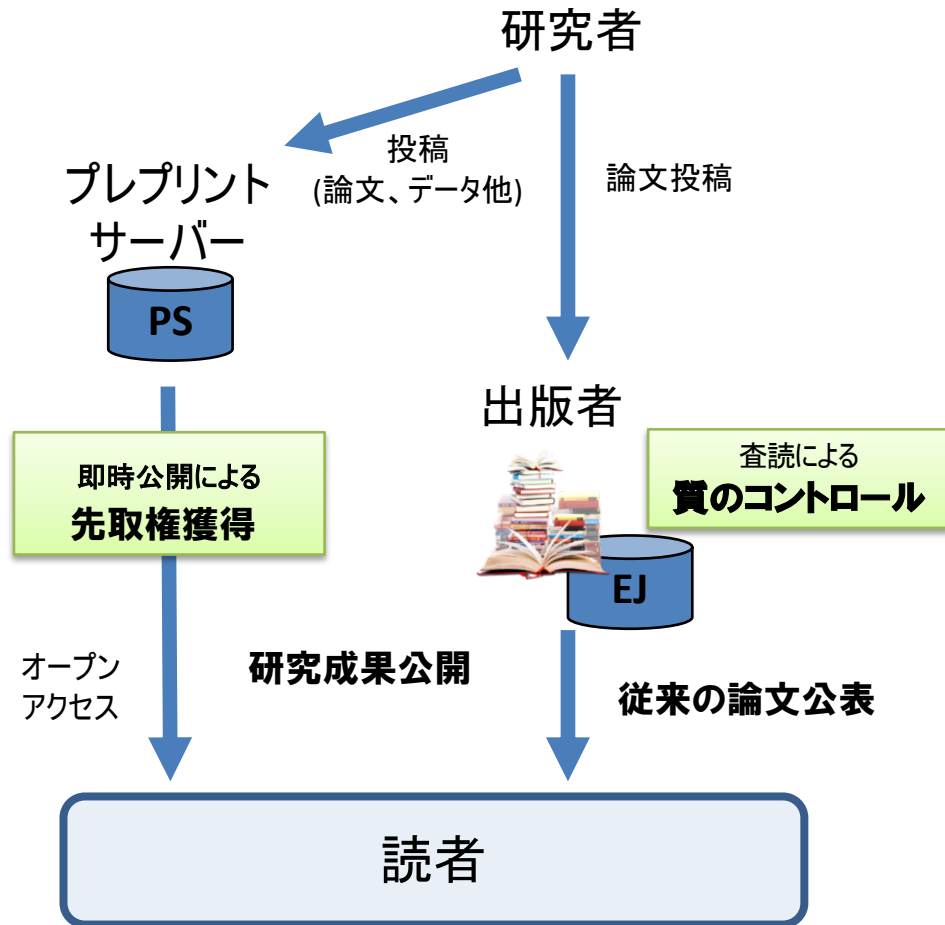
今回は，学術ジャーナルと論文の近傍のお話をします．

プレプリントサーバーの進展

従来の仕組み



プレプリントサーバ (PS) の活用



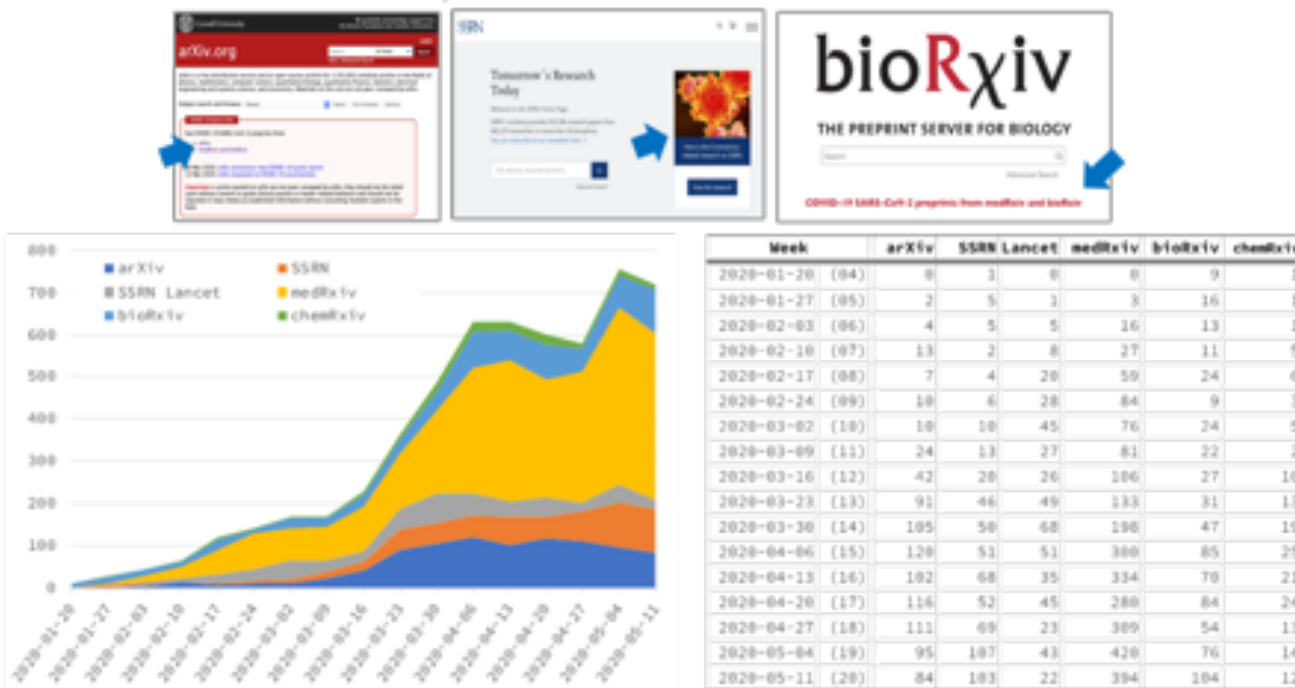
先取権を確保しつつ、査読を経て学術ジャーナルでも公開
領域によってはプレプリントサーバがメインの公開先になるところも

プレプリントによって見える研究活動の新しい“景色”

■ 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



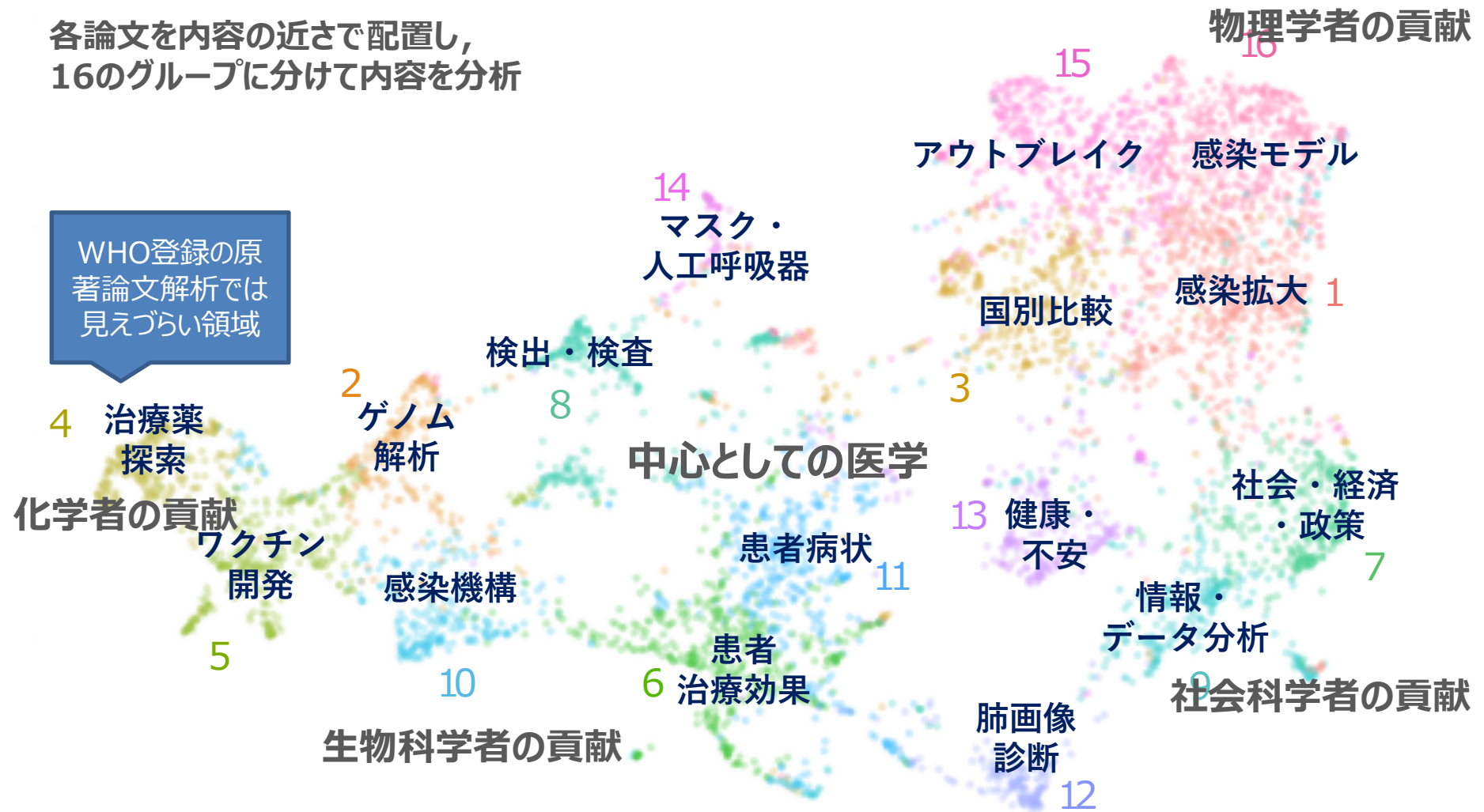
コロナ禍にあって，医学・生物・化学系だけでなく，情報，人社系など多くの分野でプレプリントが活発に投稿されている

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)

COVID-19プレプリントの分布：トピック別

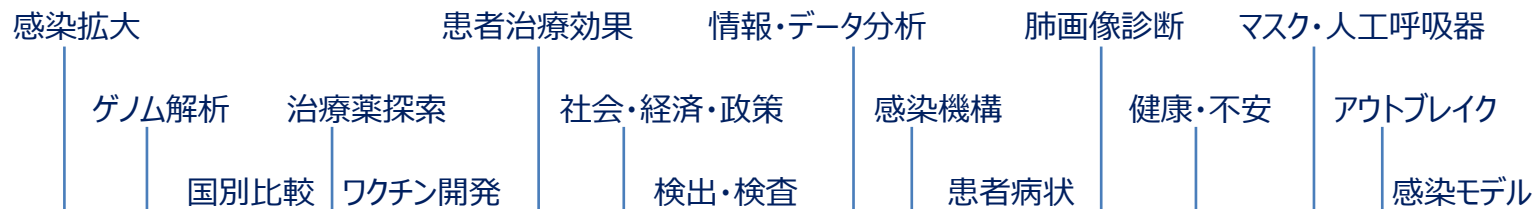
各論文を内容の近さで配置し、
16のグループに分けて内容を分析

WHO登録の原
著論文解析では
見えづらい領域



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

プレプリントサーバの種類とCOVID-19トピック



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%

話題の分布：PPS別



■ 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）

- 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)

- 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

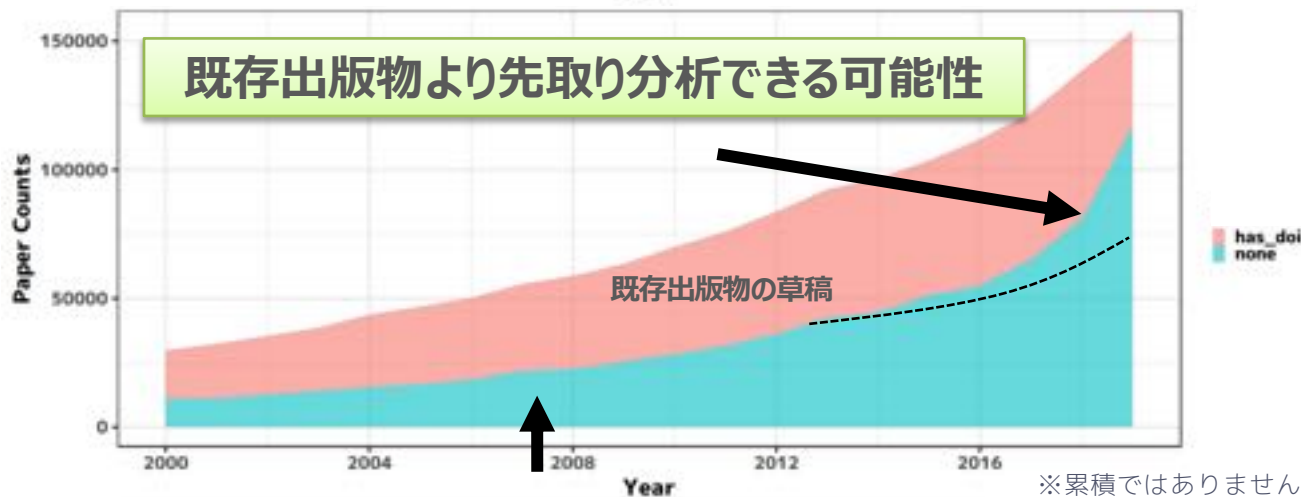
原著論文だけでは見えない分野の動向把握



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）

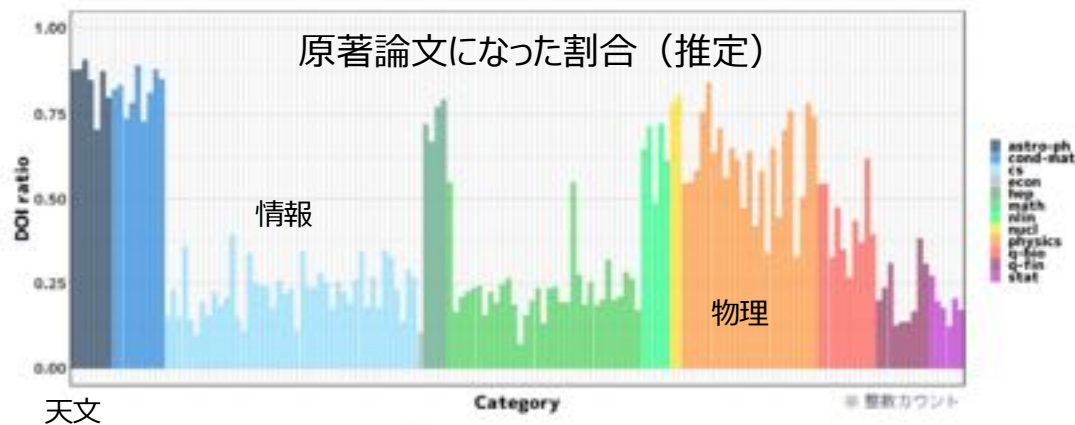


図5 分野毎のDOI付与率

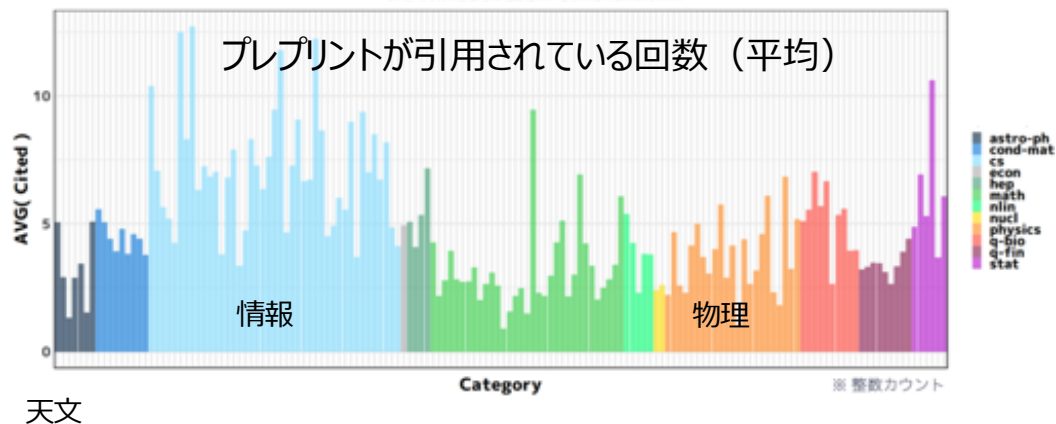


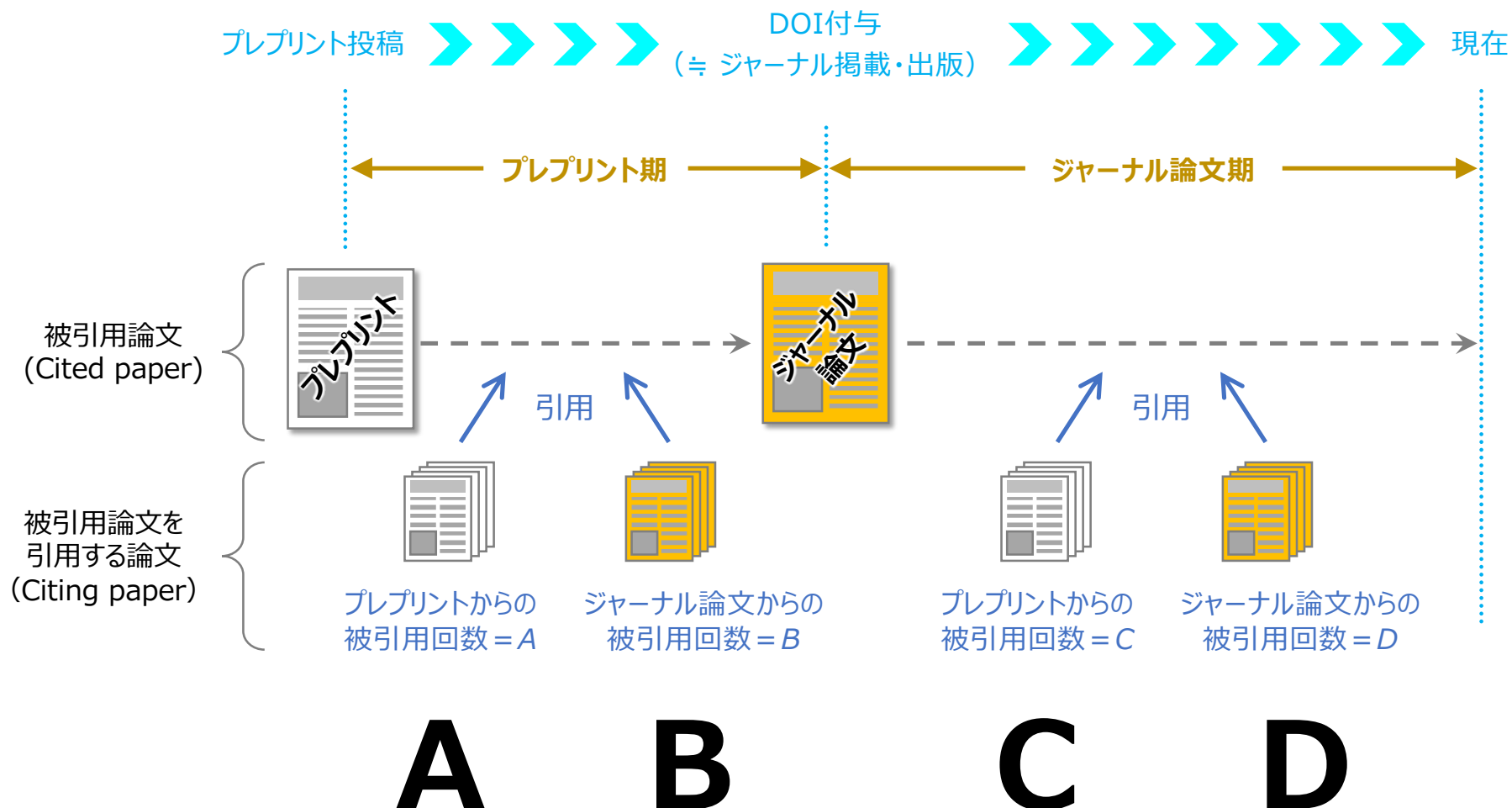
図7 分野と被引用の関係性

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

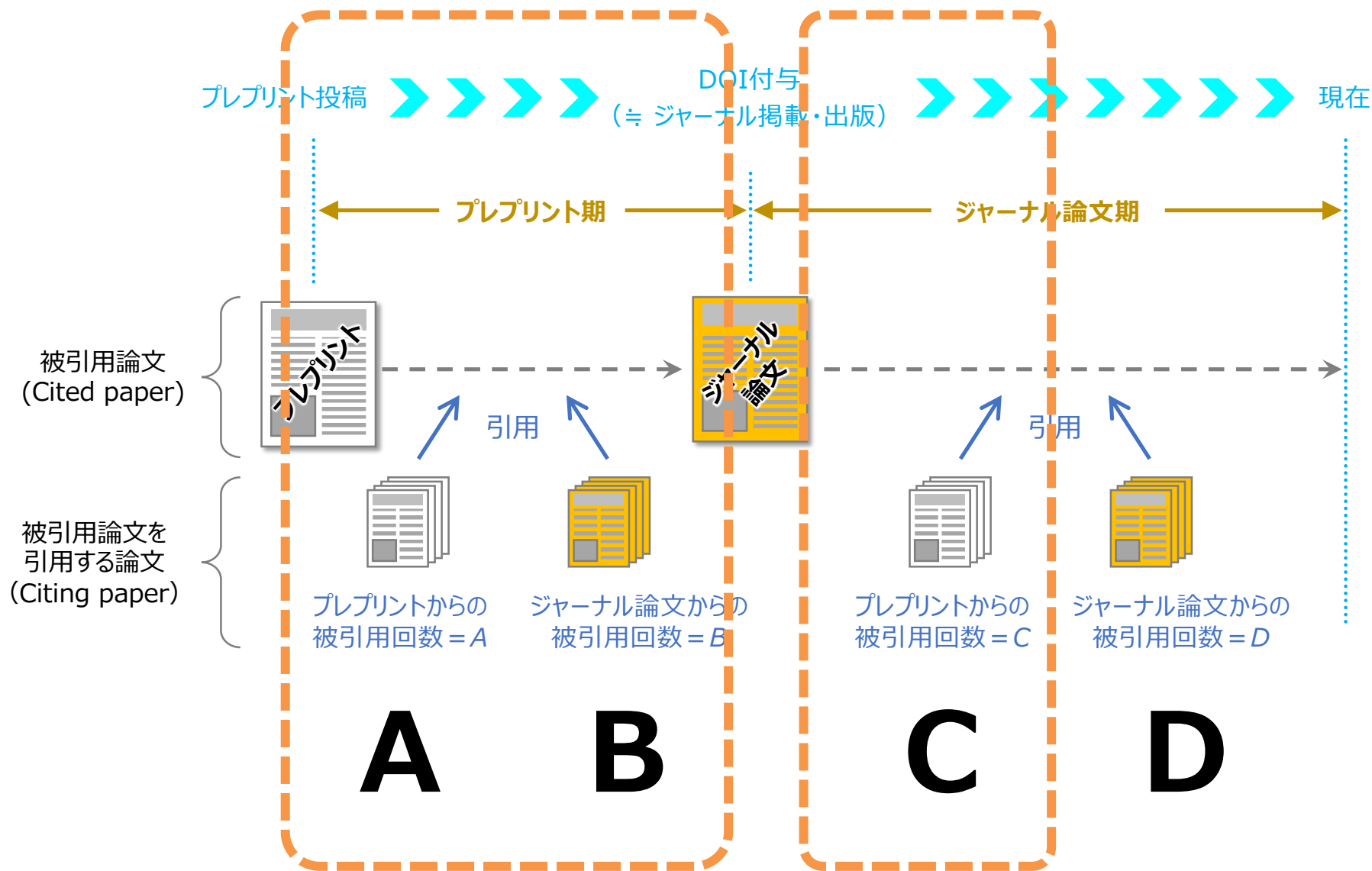
arXiv : 2020年1月21日時点で収集可能なものを全収集

データ総数： 1,622,763件 (1986年4月25日※～2020年1月17日)



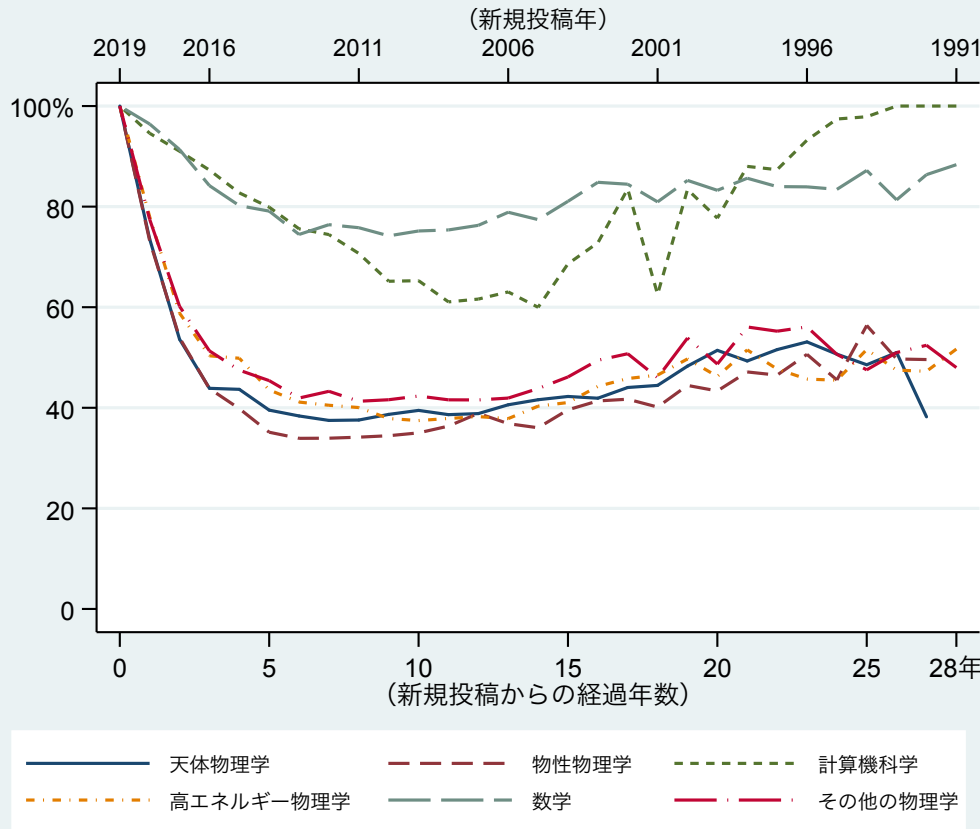
※ arXiv は 1991年スタートだが、投稿日がそれより前のモノも存在

指標の構造：プレプリントのインパクト



$$(A + B + C) / (A + B + C + D) = \text{プレプリントのインパクトの割合}$$

総被引用回数に占めるプレプリントの関与する被引用回数の割合



出典: arXiv API (<https://arxiv.org/help/api/>)及びSemantic Scholar API (<https://api.semanticscholar.org/>)からの取得データを基に著者作成。
分野カテゴリーはarXivにおける分野分類(https://arxiv.org/help/stats/2019_by_area/)に基づく。

引用活動のうち、 4割はプレプリント

- ◆ 知の生産活動のうち、引用の4割はプレプリ
- ◆ 計算機科学, 数学は6割超



研究インフラとして
不可欠な存在となっている
可能性

この調査分析報告のプレプリント版！を公開中

https://www.mext.go.jp/content/20201026-mxt_jyohoka01-000010684_2.pdf

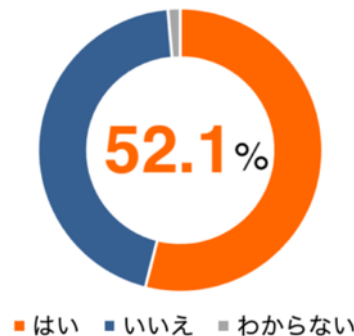
■ 調査対象

- 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%）

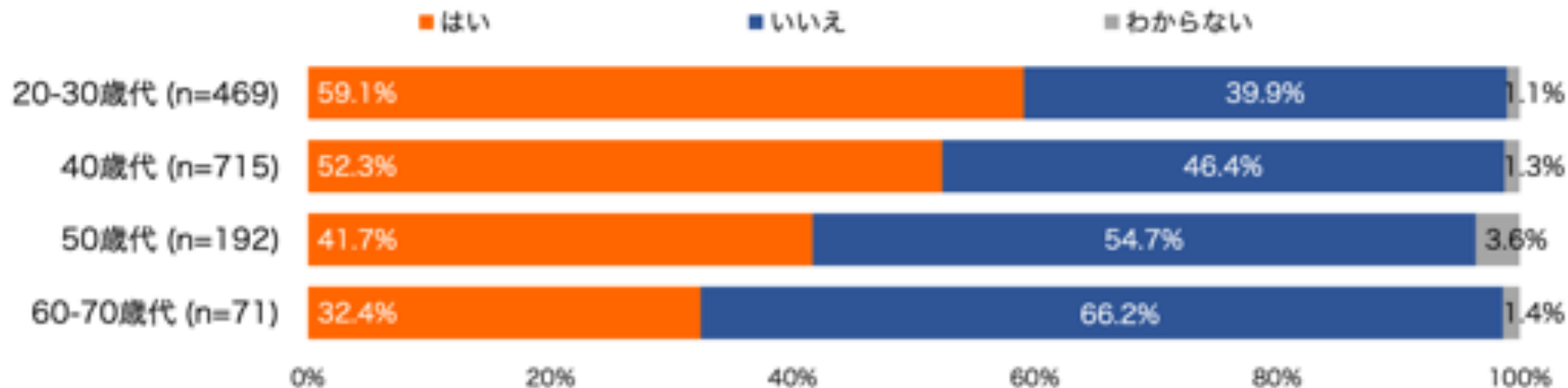
■ 調査期間

- 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

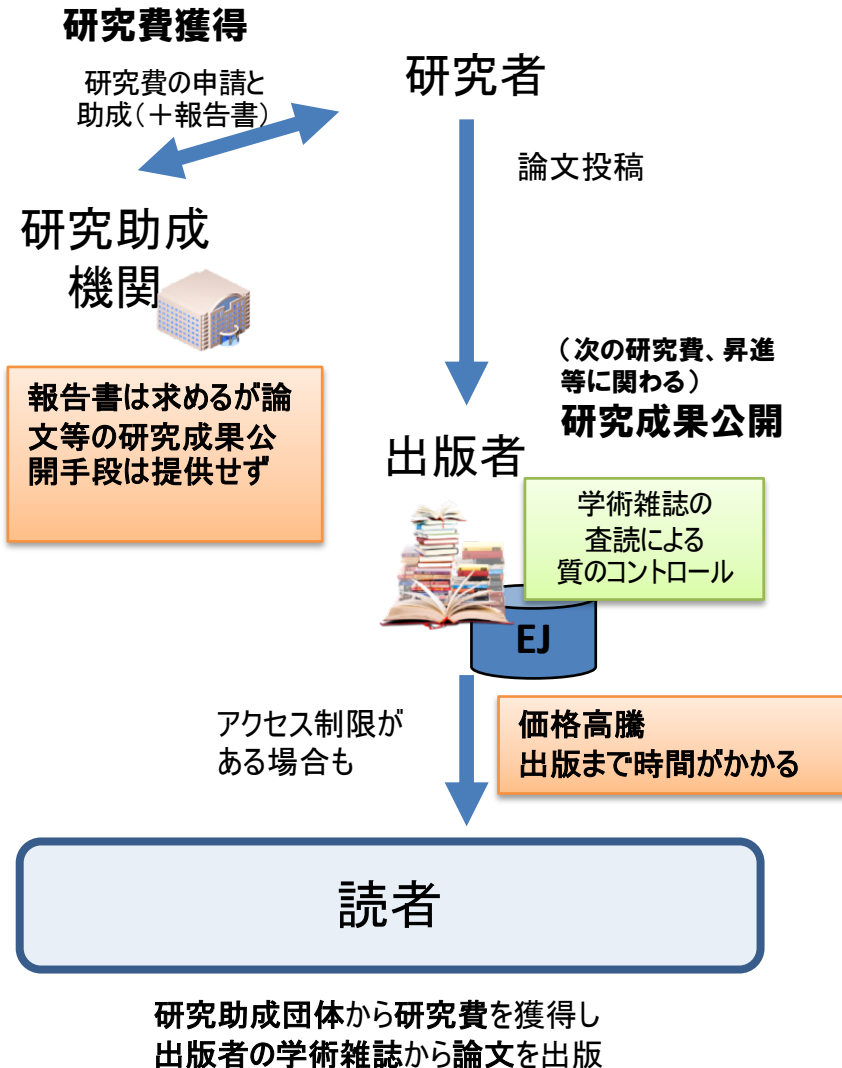


世代交代による変容を示唆

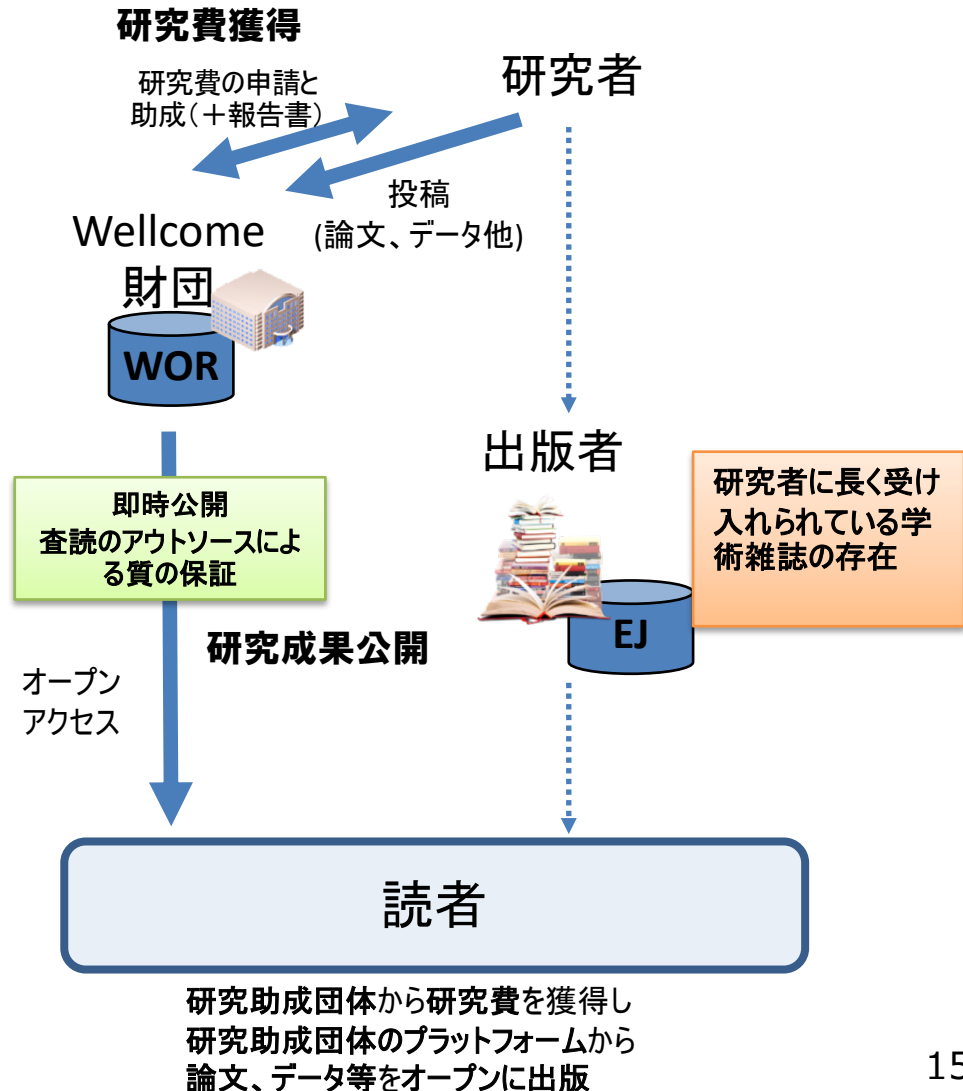
(n=1,447) ※不明(1名)を除く 14

※近日報告書発行予定

従来の仕組み



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



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

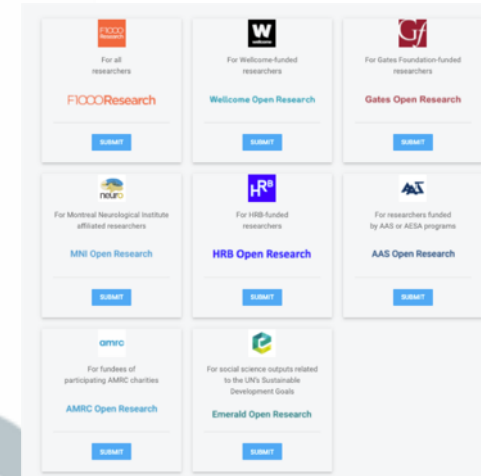
Expert peer reviewers are selected and invited, and their reports and names are published alongside the article, together with the authors' responses and comments from registered users

Publication & Data Deposition

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

Article Revision

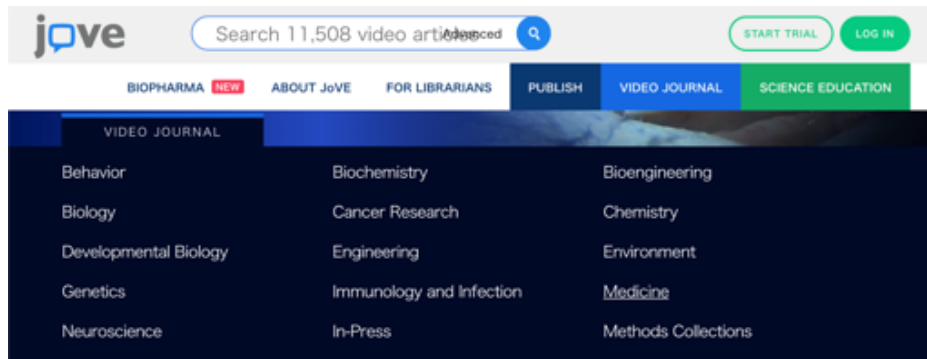
Authors are encouraged to publish revised versions of their article. All versions of an article are linked and independently citable



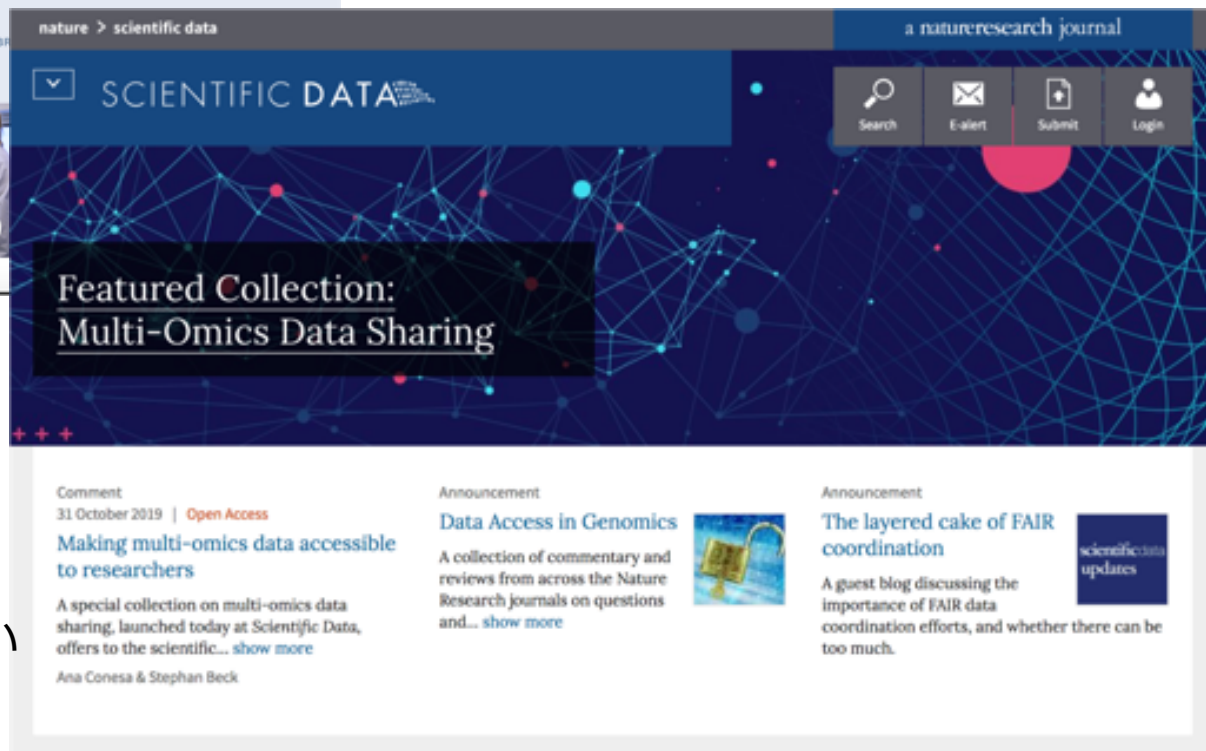
- F1000Research (筑波大学)
- ウェルカムトラスト
- ゲイツ財団
- EUで導入検討の動きも

- 論文を書いたら、データと共にまずは公開（出版）
- 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

誌面をメインとしない動画ジャーナルとデータジャーナル



実験などは動画の方が便利

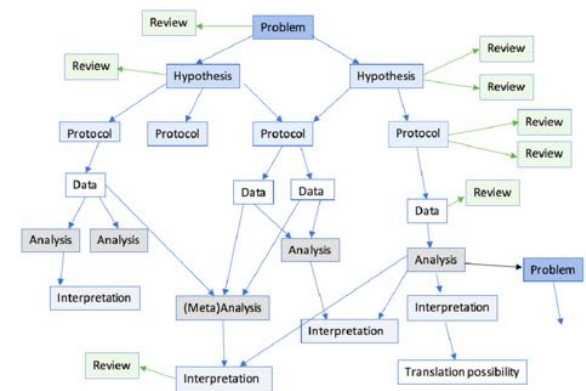


データがないと再現できない

OCTOPUS

<https://current.ndl.go.jp/node/41710>

- 英・ケンブリッジ大学のAlex Freeman氏←英・Jisc等が助成するReproducibility Networkの支援
- 研究成果の出版のプロセスを**Problems**、**Hypotheses**、**Methods/Protocols**、**Data/Results**、**Analyses**、**Interpretations**、**Applications**、**Reviews**の8個の要素に分解して公開
 - 失敗を含めた全ての研究プロセスを、研究成果の一部として公開することが可能になる。
 - 研究成果は即座に公開され、誰でもレビューと評価を与えることができる。このことによって、研究が加速する。
 - 研究者の個々の貢献を評価することが可能となる。
 - 再現性の危機に対する解決策となる。
 - 研究者間でのコラボレーションの機会を創出する。例えば、データの収集を得意とする研究者とデータの分析が得意な研究者のコラボレーションが生まれる。研究の各要素の専門化が進むことによって、研究の品質が管理される。



Cambridge scientist 'breaks up the old-fashioned academic paper'

<https://www.researchinformation.info/analysis-opinion/cambridge-scientist-breaks-old-fashioned-academic-paper>
2020年6月23日

■ 研究成果の信頼性確保とスピードのバランスをどうとるか

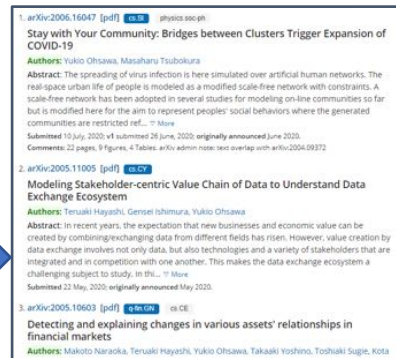
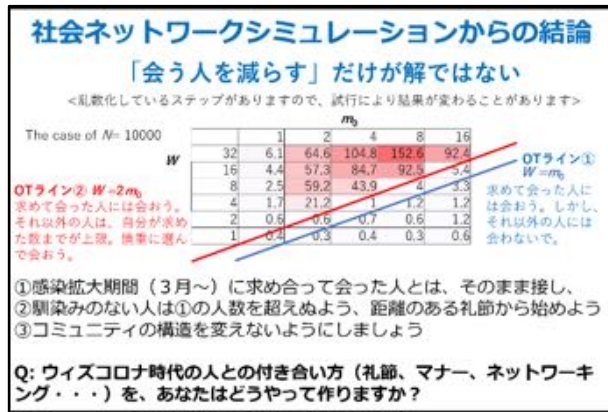
	従来の研究スタイル	新たな研究スタイル
研究の進め方	仮説・実証型	データ主導型
成果の公開方法	査読付き論文	プレプリント・研究データ
成果の価格	高価格化（ジャーナル購読料の高騰）	無料・低価格
成果公開までのスピード	査読～公開までの長いタイムラグ	速やかに公開（査読が無いため）
生まれる成果の量	少数の成果	大量の成果
公開される成果の信頼性	査読に基づく高い信頼性	質や信頼性のバラツキ増大（誤った事実やフェイクの拡散の恐れ）
スタイルの持続性	高い持続性（確立されたビジネスモデル）	不確定（未確立のビジネスモデル）
主要国	欧米日等の先進国中心	中国や新興国の躍進
研究者のインセンティブ	ハイインパクトジャーナルでの発表による高い評価	研究実績の先取権確保
有効なシーンや分野	平常時に有効	非常時（今回のコロナ対応等）に有効、技術進化の速い分野や査読に時間を有する分野に有効

内閣府研究データ基盤整備と国際展開WG（第13回）資料より抜粋

これらの変化自体はインターネット（Web）の発達とともに進行していたが、COVID-19によって大幅に加速

データからの気づきによる分野横断研究の進展

- システム創成学（チャンス発見学、データ市場創成）の研究者によるCOVID-19分析



SNSの活用

普段の研究の投稿先とは違う
 arXiv（プレプリントサーバー）
 に論文を投稿して即時公開し、
 SNSも活用して幅広い意見や
 パートナーを募る

科学
インパクト

社会
インパクト



医学、経営系情報学者等と国際コラボ*



データの再利用が進むことで興味関心を持った研究者が分野を超えて自由に研究を進めて様々な価値を発見する。←オープンサイエンスの予察が現実

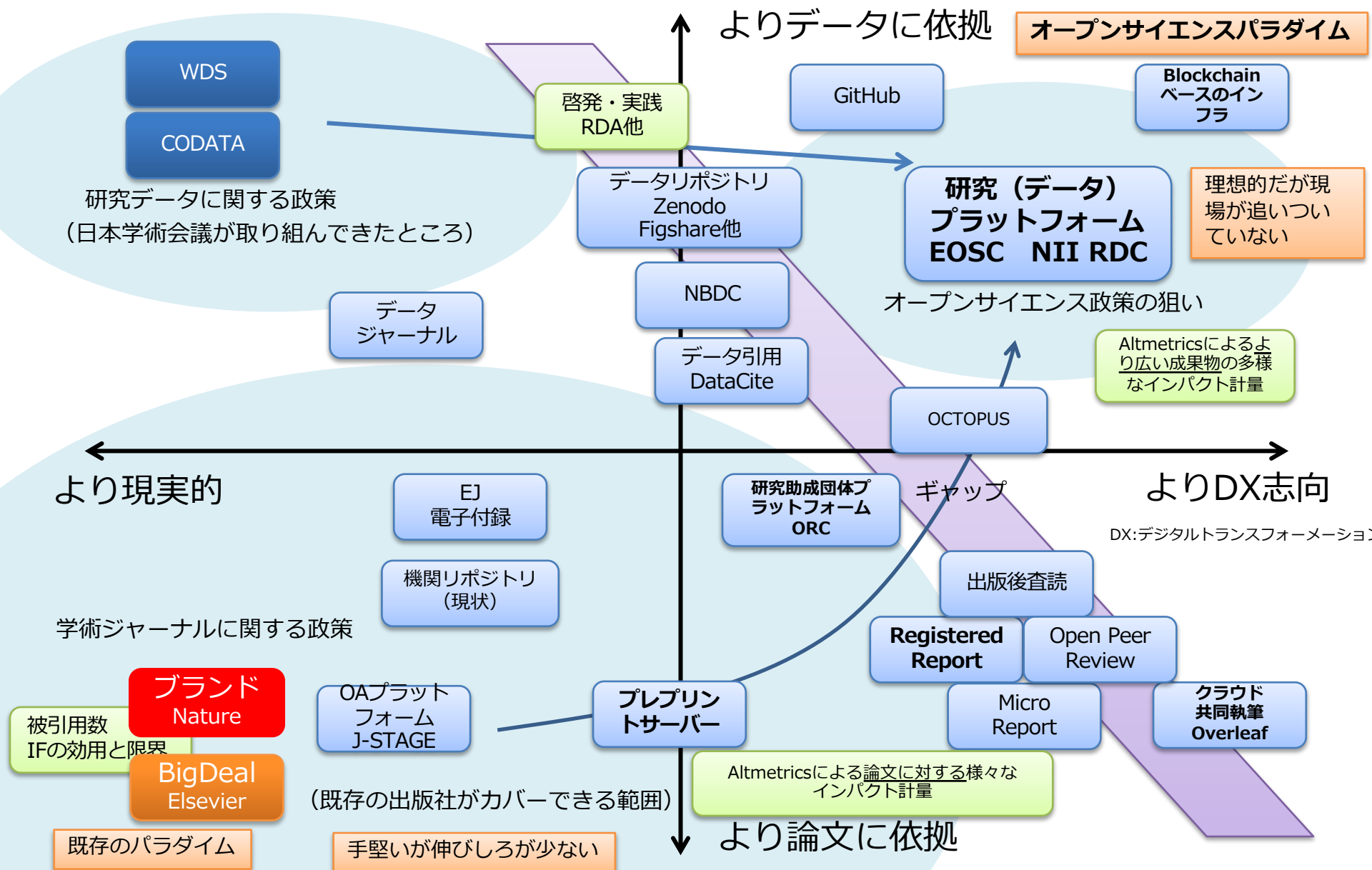
この事例では、出版者、図書館、学会が、ほぼ関与していない

* 1) U...
 2) An urgent international interdisciplinary Project "Conquer Pandemics on Constrained Social Network Models" <http://www.panda.sys.t.u-tokyo.ac.jp/covid19challengers.html>

1) ハマトーク緊急特番: <https://otagaihama.localgood.yokohama/topics/1756/>

2) 横浜市共創ラボ「Stay Home から Stay with your communityへ」 <https://www.facebook.com/LOCALGOODYOKOHAMA/videos/212777736728528> 大澤幸生先生提供

学術情報流通のDXに向けた俯瞰の例



■ 学術ジャーナルと論文のデジタルトランスフォーメーション

◆ 成果公開メディアの変容：プレプリントの可能性

- ✓ 研究者の成果公開に新たな戦術が加わった

◆ アクターの変容：新たなアクターと、出版者等が直接関与しないフロー

- ✓ 研究助成団体
- ✓ （今回は割愛したが）市民，プロボノの関わり

◆ “出版”という意味の再考：研究プロセスの各段階の共有と出版

- ✓ 研究データの公開を含むオープンサイエンスの潮流をふまえると、学術ジャーナルと論文の生態系だけを見ていると見誤る