

プレプリントをめぐる近年の動向及び今後の科学技術行政への示唆

MEXT-NISTEP プレプリント調査・検討チーム*

要旨

近年のオープンサイエンスの潮流にあつて、研究成果の発信・共有や研究コミュニティ内での評価に係る営みの顕著な割合を、研究分野によってはジャーナル論文だけでなく、その査読前段階のプレプリントが担うようになっている。急速な ICT の進展と今般の新型コロナウイルス感染症（COVID-19）流行下にあつて、こうした動きには今年（2020年）に入ってからさらに拍車がかかっている。しかしながら、研究活動におけるプレプリントの役割や位置付け、そして頓に高まるその存在感に関する定量的なエビデンスは、これまで研究者や政策関係者の間でもごく限定的にしか語られてこなかった。こうしたプレプリントをめぐる近年の動向や研究現場の実態を踏まえ、今後の学術情報流通の在り方についても一定の見通しを持ちつつ時宜を得た科学技術政策へとつなげていくため、本稿ではまずプレプリントの利点・有用性やリスク等について現代的な視点から改めて整理しつつ、プレプリントにまつわる「質」評価の在り方について再考する。その上で、プレプリントサーバーの先駆けである arXiv 上の大規模データを分析することで、現代の知の生産活動におけるプレプリントの意義と役割について試行的な検証を行う。今回の計量書誌学的アプローチに基づく分析の結果からは、論文引用を通じた知の生産活動の顕著な割合がプレプリントの仕組みに大きく依存している実態が研究分野の特性とともに浮かび上がる。あわせて、オープンサイエンス時代の研究評価の在り方等の観点を含め、プレプリントを通じた研究進展がもたらす今後の科学技術政策上の示唆について議論する。

キーワード： プレプリント、オープンサイエンス、研究評価、科学技術政策、EBPM、計量書誌学

* 岡村 圭祐（MEXT）、依田 洸（MEXT）、林和弘（NISTEP）、小柴 等（NISTEP）

目次

1	プレプリントとは（概観）	1
1.1	はじめに	1
1.2	プレプリントの「質」	3
1.3	プレプリントの利点・有用性	6
1.4	プレプリントのリスク・課題	7
1.5	情報科学系分野の特性	9
2	知の生産活動におけるプレプリントの役割（定量的分析）	10
2.1	分析手法	10
2.2	データ	12
2.3	分析結果及び考察	13
3	今後の科学技術行政への示唆	17
3.1	研究評価の観点	17
3.2	科学コミュニケーションの観点	19
3.3	日本の役割	20
3.4	おわりに	22
	参考文献	23

図目次

1	プレプリントの仕組み	2
2	bioRxiv に掲載された COVID-19 関連のプレプリント	5
3	arXiv ウェブサイト（トップページ）	11
4	被引用回数のカウントに関する整理	12
5	「プレプリント期」の長さ	13
6	arXiv への新規プレプリント投稿数	14
7	arXiv 上の論文が新規に獲得した被引用回数	14
8	総被引用回数に占めるプレプリント期に獲得した被引用回数の割合（‘ α 値’）	15
9	総被引用回数に占めるプレプリントから獲得した被引用回数の割合（‘ β 値’）	15
10	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（‘ γ 値’）	16
11	「arXiv 版インパクトファクター」	16

表目次

1	主なプレプリントサーバー一覧（2020 年 10 月現在）	4
2	6 分野カテゴリー（arXiv 分類の大括り化）	13

1 プレプリントとは（概観）

本節ではまずプレプリントの性質や特徴とその歴史的経緯等について概観する。その上で、プレプリントの利点・有用性やリスク等について現代的な視点から改めて問い直し、今般のコロナ禍にあっての社会的な受け止めも踏まえ、プレプリントの「質」評価の在り方について再考する。

1.1 はじめに

研究者は他の研究者との交流にあたって研究論文の執筆・発表を行う [1]。多くの場合、研究論文の主な発表舞台である学術ジャーナル（論文誌、ジャーナル誌）には、その出版プロセスにおいて、各学問分野に精通した研究者や専門家が事前に論文内容の評価を行う査読（ピアレビュー）と呼ばれるプロセスが組み込まれており、これはジャーナル側にとっては質担保の観点からのスクリーニングとして機能する。ジャーナル編集部は査読者（レビュア、レフェリー）のコメントを踏まえて論文の著者に対して論文の不十分な点についての改善を求め、著者はこれに逐次対応していくことが求められる。したがって論文投稿から最終的な掲載・出版までに場合によっては年単位の時間を要することも珍しくない [2, 3]。このように査読付きジャーナル論文では投稿から出版までの期間が比較的長くなるが、そうしたジャーナル側での査読を受けていないステータスで、著者による投稿と同時に近いタイミングで公開される論文原稿が「プレプリント」である。プレプリントは、DOI¹あるいはそれに準じた独自の識別子（ID）が付与されることで広くデジタル世界で認知されるメリットに加え、一般的なユーザーは金銭的な負担なく投稿・閲覧が可能なることもあり、近年そのユーザー層は急激に拡大してきた。プレプリントを通じた研究成果の共有様式（“プレプリント様式”：図 1）は、近年のオープンサイエンスの潮流にあって、いまや多くの研究分野において研究活動やその評価の在り方を考える上で欠かせない要素となっている [4, 5]。

そうしたプレプリントを公開・管理するサーバーであるプレプリントサーバーの先駆けとなったのが、1991年8月に運用開始された arXiv [6, 7] である²。後述のとおり、ライフサイエンス系や医学系の分野でプレプリント様式が取り入れられるようになったのは 2010 年代に入ってからのことだが、物理学、数学、計算機科学等のいわゆる数物系の分野では arXiv が既にこの 30 年近くの間、研究コミュニティの活動にとって欠かせないプラットフォームを提供してきた。先行研究や最新の研究動向の把握、参考文献の収集、論文の執筆、ジャーナル掲載に先立つ先行的な公開・成果共有、成果に関する先取権の獲得、研究コミュニティからのフィードバックを踏まえた出版前の随時アップデートに至るまで、基本的な研究活動のおおよその部分がプレプリントを通じた研究成果の共有様式で十分に完結するとの指摘もあるほど、分野によってはその活用が進んできた。一部のプレプリントサーバーでは、原稿そのものに加えてデータやプログラムのソースコードや手法までを含めた多様な種類・様式での成果共有も可能であり、プレプリントサーバーから査読付きジャーナルへの直接投稿を受け付けるなど、ジャーナル側でもプレプリント・フレンドリーな仕組みが次々と導入されてきた³。

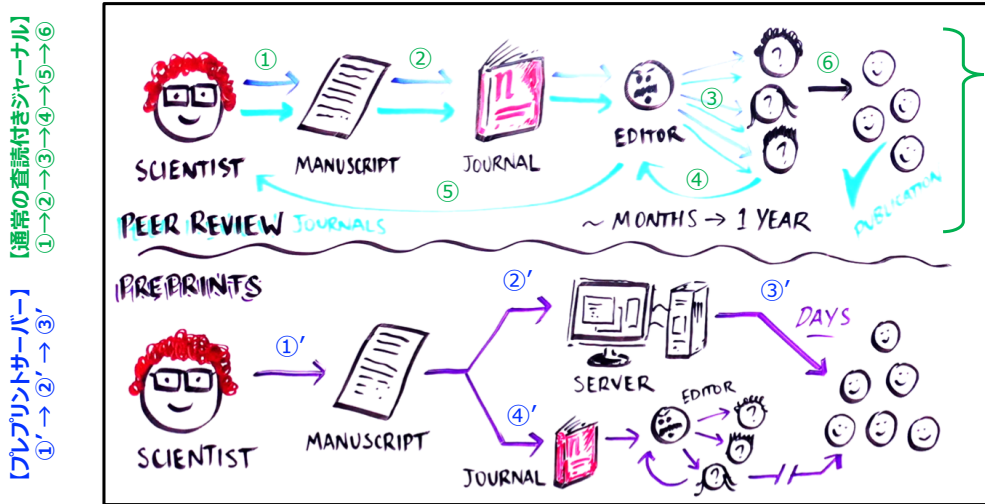
このように存在感を増すプレプリントの意義や役割、そして分野特性を踏まえた留意点等については多くの識者により語られてきた（簡易な読み物としては [8, 9] 等）。一般論として、ジャーナル論文とプレプリントとは、その公開までの期間の長短や論文の「質」保証等の観点から互いに相補的な機能が期待されるものであり、その相補性の程度も現状では分野によって大きく異なるのが現状である。例えば物理学分野では、多くの場合、論文著者は、その原稿をまずプレプリントとして arXiv に投稿しつつ、同時に（あるいは少し時間をおいて）ジャーナルにも平行して投稿することで、最終的には両方のメリットを享受しようとすることが多い。そうした研究者の行動原理を踏まえ、いまや多くのジャーナルがプレプリントを経ての論文投稿やその論文中のプレプリントの引用を認めている [10, 11]。

¹ Digital Object Identifier：インターネット上のドキュメントに付与される永続的デジタル識別子。

² ‘arXiv’ は英単語 ‘archive’ と同じく「アーカイブ」と発音する（‘X’ をギリシャ文字「カイ (chi; χ)」の大文字と見なしている）。

³ 一部の物理系分野ではプレプリント ID の通知のみでジャーナルへの論文投稿が完了する仕組みも既に導入されている。

- ① 研究者は論文原稿を書く。
- ② それをジャーナルに投稿する。
- ③ もしジャーナルのエディタが関心を示せば、その論文原稿は匿名の査読者に送られ、ピアレビュー・プロセスに入る。
- ④ ジャーナル側は査読者からのコメントを踏まえ、その論文をジャーナルに掲載（受理）すべきか却下かを判断する。
- ⑤ 却下されなければ、多くの場合エディタは論文著者に対して査読者コメントを踏まえた論文の書き直しを求め、それをもとに査読者から再ピアレビューがなされる。
- ⑥ その上で最終的な採否判断がなされる。

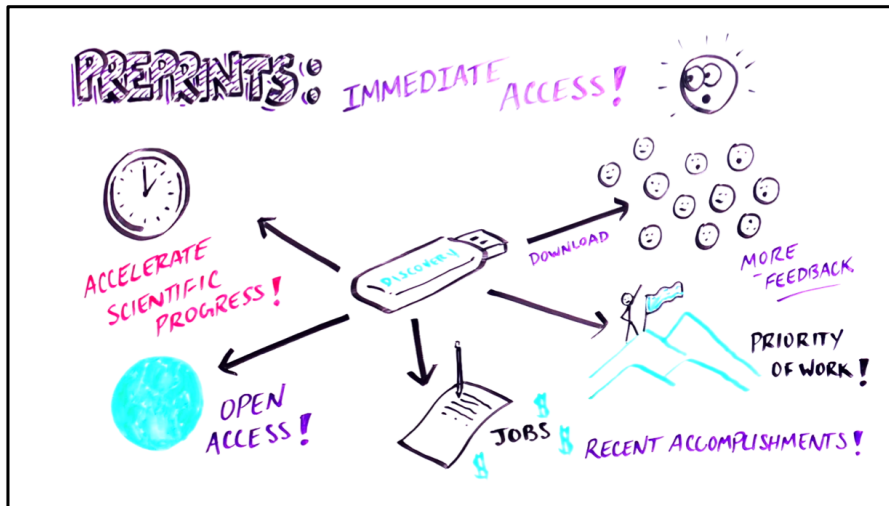


⑦ このプロセスは一定程度機能しているが時間がかかる。また、優れた研究成果であっても査読者次第で却下されてしまうことがある。

- ①' 研究者は論文原稿を書く。
- ②' それをプレプリントサーバーにアップロードする。
- ③' プレプリントは学術論文としての最低限のチェックを経た上で、速やかに研究コミュニティ全体に対してオンライン公開される。
- ④' 通常、その論文原稿はそのまま査読付きジャーナルへも並行して投稿され、上の③～⑥のプロセスを経る。

(プレプリントへの即時アクセス可能な状況はいくつものアドバンテージが伴う)

▶ 新しい研究成果へのアクセスを迅速化することによって分野の研究進展を加速させることができる。



◀ 匿名の数名の査読者から査読される場合に比べ、より幅広い層からフィードバックを得ることができる。

◀ 研究成果についての先取権・クレジットを即座に宣言・獲得できる。

▲ プレプリントはオープンアクセスであり、インターネット環境があれば世界中の誰もがその論文原稿を無料で閲覧することができる。

▲ 進行中の研究プロジェクトやつい最近の（したがってまだジャーナルには掲載されていない）研究成果について、ファンディング・エージェンシーや現在・将来の雇用主に対して報告・アピールすることができる。

【図 1】 プレプリントの仕組み (YouTube 動画 [12] をもとに著者にて枠外に説明を追記)

こうした中、プレプリント様式は近年、その対象分野が大きく拡大してきている [13, 14]。ライフサイエンス分野の bioRxiv [15] (2013 年 11 月開始)、化学分野の ChemRxiv [16] (2017 年 8 月開始)、医学分野の medRxiv [17] (2019

年6月開始)をはじめ、プレプリントを通じた研究成果の公開は分野特化型あるいは分野横断型の様々な形態で導入が進み、この数年でも数多くのプレプリントサーバーが立ち上がってきた [18, 19]. 表1はこれまでに整備されてきた主なプレプリントサーバー(プレプリントリポジトリを含む)を一覧で整理したものであるが、本表からもプレプリントを通じた研究様式がいまやあらゆる分野(人文科学系や社会科学系の分野も含めて)に浸透しつつあることが見て取れる。このうち特に医学系やバイオ系のプレプリントは、昨今の新型コロナウイルス感染症(COVID-19)の感染拡大抑制や治療法確立に向けた研究論文のオープン化や公表迅速化の流れを受け、研究者の研究活動に特に大きな影響と変化をもたらしてきた。実際、プレプリントという様式がこれまでになく社会的な関心までも集めるようになったのは、今般のコロナ禍にあつてのCOVID-19に関する情報発信が一つの大きな契機となっている。いまこの瞬間もそうであるように、世界中の研究者がCOVID-19研究に打ち込む間にも感染者は増え続けている。そのような中、最新の実験結果や研究成果、データや科学的知見がより早く発信され共有されるほど、世界中の研究者がそうした情報により早くアクセスでき、自身の研究にも速やかに活用・反映できるため、より早い治療法の確立につながる可能性が高まる。この時間との闘いの中で情報の発信や共有に際して重要なプラットフォームを提供してきたのがプレプリントサーバーであることは疑いない。実際、2019年に運用開始したばかりの医学系のプレプリントサーバーであるmedRxiv [17]では、COVID-19に関連するものに限定しても、2020年1月中旬から5月上旬までの約5か月間で3千件近くのプレプリント投稿が行われている [20].

こうしたケースとともに「プレプリント」という研究成果の共有様式が社会的にも広く認知されていく中で、情報の信頼性や伝達の正確性といった観点で、これまでにない学術情報と社会との関わり方の難しさが浮き彫りとなってきたことも事実である [21]. 今後、プレプリント様式が研究の世界のみならず社会全体に及ぼしていく可能性のある影響やその生み出していく変化とはどのようなものか。その上で、それを先取りした上で今後の科学技術・イノベーション行政はいかにあるべきか。こうした問題意識の下、今般、文部科学省と科学技術・学術政策研究所(NISTEP)の有志チームにおいて、実際のプレプリントサーバーのデータをもとにした定量的なエビデンス収集・分析とあわせて、プレプリントをめぐる各種の動向や経緯等について見識の深い有識者の方々 [22–24]からもヒアリングを行うことで、プレプリントをめぐる諸課題に機動的に向き合いつつ時宜を得た政策実装へとつなげていくための検討を行ってきた(2020年8月下旬~9月)。本稿ではその成果について報告する。

1.2 プレプリントの「質」

査読付きジャーナルでは論文投稿から公開までに一定程度(場合によっては年単位で)の期間を要するが、ジャーナル側での査読を経ていないプレプリントの初稿であれば、多くの場合当日または数日以内にはプレプリントサーバー上で一般に公開される。結果として、正しく意義深い研究成果が速やかに公表され、アカデミアや社会に新たな価値を生んでいくのであれば、誰にとっても望ましいことであるが、他方で、拙速に発表されようとする研究成果の中には必ずしも質の伴わないものがあることも事実である。昨今、プレプリントについて、ジャーナルによる査読を経ていないがために質が低いとの言説が行政の周辺でも多く聴かれるが、その背景として多くの人の印象に残るものは、新型コロナウイルス感染症(COVID-19)とHIVとの類似性を主張したbioRxiv上のプレプリント [25]が(論文に書かれたそのままの内容ではなく)誤った形で情報流通し、その後即撤回された最近の事例であろう。図2に当該プレプリントの掲載されていたウェブページの一部画面を示すが、下部のオルトメトリクスや公の閲覧者からのコメント数の多さからも見て取れるとおり、この事例は研究コミュニティのみならず社会的にも大きな関心を引き起こし、これは従来の査読付きジャーナルでは見られなかった社会とアカデミアとの間の科学技術コミュニケーション上の新展開を象徴するものとも言える。ただし、これもよく知られるとおり、COVID-19関係での撤回事例は何もプレプリントに限ったことではなく、権威ある査読付きジャーナル誌においても同様に見られている。例えば、著名な医学誌である英Lancet誌と米New England Journal of Medicine (NEJM)誌でも、COVID-19に関して掲載された論文が本年6月初めに相次いで撤回されるという事態が生じており、コロナ禍にあつて論文の査読が甘くなっていることを憂慮する声とともに国内外で多くの報道がなされたことは記憶に新しい。COVID-19関連研究に限らずとも、このよう

【表 1】 主なプレプリントサーバー一覧 (2020 年 10 月現在)

プレプリントサーバー	開始年	対象分野	識別子	運営・開発元
arXiv	1991	物理学, 数学, 計算機科学等	独自 ID	コーネル大学図書館 [米]
SSRN	1994	全般 (当初は社会科学)	独自 ID	エルゼビア社 [蘭]
Cryptology ePrint Archive	1996	暗号学	独自 ID	国際暗号学会 (IACR)
CogPrints	1997	認知科学, 心理学, 神経科学等	独自 ID	サウサンプトン大学電子工学計算機科学科 [英]
RePEc	1997	経済学	-	世界中の研究コミュニティ有志
LingBuzz	2001	言語学	独自 ID	トロムソ大学 [ノルウェー]
PhilSci Archive	2001	科学哲学	独自 ID	ピッツバーグ大学哲学科 [米]
e-LiS	2003	図書館情報学	-	フェデリコ 2 世 ナポリ大学 [伊]
Nature Precedings	2007	全般	DOI	シュプリンガー・ネイチャー社 [英]
OSF Preprints	2007	※プレプリントサーバーの集約機能	DOI	オープンサイエンス フレームワーク (OSF)
ResearchGate	2008	全般	DOI	ResearchGate 社 [独]
PhilArchive	2009	哲学	独自 ID	ウェスタンオンタリオ大学デジタル哲学センター [加]
viXra	2009	arXiv の各分野に加えて人文科学等	独自 ID	Philip Gibbs
Authorea	2012	全般	DOI	Atypon 社 [米]
bioRxiv	2013	生物学	DOI	コールドスプリングハーバー研究所 [米]
PeerJ Preprints	2013	生物学, 生命科学, 医学	DOI	PeerJ [米]
Zenodo	2013	全般	DOI	欧州原子核研究機構 (CERN)
JMIR Preprints	2015	医学, e-Health	DOI	JMIR Publications 社 [加]
ChinaXiv*	2016	中国研究コミュニティ向け	独自 ID	中国科学院
engrxiv	2016	工学	DOI	オープンサイエンス フレームワーク (OSF)
Preprints.org	2016	全般	DOI	MDPI 社 [スイス]
PsyArXiv	2016	心理学	DOI	オープンサイエンス フレームワーク (OSF)
SocArXiv	2016	社会科学	DOI	オープンサイエンス フレームワーク (OSF)
agriRxiv	2017	農学	DOI	オープンサイエンス フレームワーク (OSF)
ChemRxiv	2017	化学	DOI	米英独中の 5 化学会
EarthArXiv	2017	地球科学	DOI	オープンサイエンス フレームワーク (OSF)
ESSOAr	2017	地球宇宙科学	DOI	米国地球物理学連合, Atypon 社 [米]
FocUS Archive	2017	超音波研究	DOI	オープンサイエンス フレームワーク (OSF)
INA-Rxiv*†	2017	インドネシア研究コミュニティ向け	DOI	オープンサイエンス フレームワーク (OSF)
LawArxiv	2017	法学	DOI	オープンサイエンス フレームワーク (OSF)
LISSA	2017	図書館情報学	DOI	オープンサイエンス フレームワーク (OSF)
MarXiv	2017	海洋科学	DOI	Zenode (CERN)
MetaArXiv	2017	研究公正	DOI	オープンサイエンス フレームワーク (OSF)
MindRxiv	2017	心と瞑想の科学	DOI	オープンサイエンス フレームワーク (OSF)
NutriXiv	2017	栄養学	DOI	オープンサイエンス フレームワーク (OSF)
SportRxiv	2017	スポーツ学	DOI	オープンサイエンス フレームワーク (OSF)
PaleorXiv	2017	古生物学	DOI	オープンサイエンス フレームワーク (OSF)
Therapoid	2017	治療学	-	Open Therapeutics 社 [米]
Thesis Commons	2017	学位論文 (分野を問わない)	DOI	オープンサイエンス フレームワーク (OSF)
advance	2018	全般	DOI	SAGE 社 [米]
AfricArXiv*	2018	アフリカ研究コミュニティ向け	DOI	オープンサイエンス フレームワーク (OSF)
Arabixiv*	2018	アラブ研究コミュニティ向け	DOI	オープンサイエンス フレームワーク (OSF)
ECSArXiv	2018	全般 (当初は電気化学, 固体物理学)	DOI	オープンサイエンス フレームワーク (OSF)
FrenXiv*	2018	フランス研究コミュニティ向け	DOI	オープンサイエンス フレームワーク (OSF)
Research Square	2018	全般	DOI	Research Square Company [米]
APSA Preprints	2019	政治学	DOI	ケンブリッジ大学出版 [英], 米国政治学会 (APSA)
Beilstein Archives	2019	有機化学, ナノテクノロジー	DOI	バイルシュタイン協会 [独]
BodoArXiv	2019	中世研究	DOI	オープンサイエンス フレームワーク (OSF)
EcoEvoRxiv	2019	生態学, 進化学	DOI	オープンサイエンス フレームワーク (OSF)
EdArXiv	2019	教育学	DOI	オープンサイエンス フレームワーク (OSF)
IndiaRxiv*	2019	インド研究コミュニティ向け	DOI	オープンサイエンス フレームワーク (OSF)
MediArXiv	2019	メディア学, コミュニケーション学	DOI	オープンサイエンス フレームワーク (OSF)
medRxiv	2019	臨床医学	DOI	コールドスプリングハーバー研究所 [米]
preprints.ru*	2019	ロシア研究コミュニティ向け	DOI	国立電子情報コンソーシアム (NEICON) [露]
TechRxiv	2019	工学全般	DOI	IEEE 財団 [米]
BioHackathon	2020	BioHackathon (国際開発者会議) 用	DOI	オープンサイエンス フレームワーク (OSF)
Cambridge Open Engage	2020	全般	DOI	ケンブリッジ大学出版 [英]
SciELO Preprints*	2020	ラテンアメリカ, イベリア半島, 南ア	DOI	科学電子図書館オンライン (SciELO)

註) 本表ではプレプリントサーバーとプレプリントリポジトリとを厳密に区別しない。また、これらいずれかの機能を含む学術情報プラットフォームの提供サービスの類も含む。「プレプリントサーバー」のうち「★」印を付してあるものは国・地域特化型のもの、「†」印を付してあるものは既にクローズ (閉鎖, 廃止) されているものである。「開始年」はサーバー/リポジトリが立ち上がった年または最初のコンテンツが投稿された年を表す。「識別子」はプレプリントに付与される固有の識別子を表す。「開始年」が同一の場合の掲載はアルファベット順による。

に公開までのプロセスの顕著な短縮化をもたらし得る弊害は、いまやプレプリントか査読付きジャーナル論文であるかの区別無しに同様に言えることである。

そもそもプレプリントの「質」についての議論は、今般 COVID-19 流行下で注目を集めるよりも前からアカデミア内外で議論されてきた。その中には、プレプリントは「混乱と歪み」をもたらすと主張 [26] もあれば、プレプリントがジャーナル論文と比較して必ずしも (平均的には) 質が劣っているとも言えず、両者は相補的に活用されるべきものとの主張 [27] もあり、現在でも議論が継続している。こうした研究の「質」についての定量的な検証としては、バイオ系の分野においては最近でも観測的手法に基づく先行研究があり、そこでは両者の「質」については統計的な有意差は小さく、したがってプレプリントはジャーナル論文と比較しても学術的に遜色のないものであると結論付け

The screenshot shows the bioRxiv preprint interface. At the top, there is a navigation bar with links for HOME, ABOUT, SUBMIT, NEWS & NOTES, ALERTS / RSS, and CHANNELS. A search bar is also present. Below the navigation bar, a yellow banner contains a warning: "bioRxiv is receiving many new papers on coronavirus SARS-CoV-2. A reminder: these are preliminary reports that have not been peer-reviewed. They should not be regarded as conclusive, guide clinical practice/health-related behavior, or be reported in news media as established information." The main content area features the title "Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag" and the authors' names. It includes a "WITHDRAWN" status, a "Comments (122)" link, and a "Previous" button. The article was posted on February 02, 2020. There are options to "Download PDF", "Email", "Share", and "Citation Tools". Social media sharing buttons for "Tweet" and "いいね!" (2,746) are also visible. The "Metrics" tab is selected, showing "Article lifetime" filters for "Article lifetime", "Last 6 months", and "This month". A table shows "Article usage: January 2020 to October 2020" with columns for "Abstract" (619,719) and "PDF" (1,504,753). A circular chart shows 14256 interactions, with a legend indicating: Picked up by 100 news outlets, Blogged by 23, Referenced in 1 policy documents, Tweeted by 16956, On 22 Facebook pages, On 11 videos, and 180 readers on Mendeley. On the right, there is a "Subject Area" section for "Evolutionary Biology" and a "Subject Areas" list including Animal Behavior and Cognition, Biochemistry, Bioengineering, Bioinformatics, and Biophysics.

【図 2】 bioRxiv に掲載された COVID-19 関連のプレプリント [15, 25]

られている [28] (他にも [29])。また、バイオ分野ではプレプリントを経た論文のほうが被引用回数上もオルトメトリクス上もより高い注目を集めるとの研究成果もある [30]。

ここで留意すべき点は、今般のコロナ禍にあって緊急性を要する研究成果や社会的ニーズの特に高い研究成果についての論文に対する査読が「甘くなる」ことが仮にあったとして、そのことによる質の非保証（の可能性）の問題と、一般のプレプリントについて、その初稿投稿時にジャーナル側での査読を経ていないことによる質の非保証（の可能性）の問題とは明確に区別して論じられるべきという点である。本稿で「研究の質」の観点を論じる際には、主に一般論としての後者の観点到立つ。その上で、まず以下の点を指摘しておきたい。そもそもプレプリントの質というのは、ジャーナル論文と同様、それぞれの研究コミュニティ内の価値基準ないしアカデミック・インテグリティ（研究公正）によって担保されていることが前提である。コミュニティ内での立場や信用が掛かっているため、たとえプレプリントであろうと、ジャーナル論文の場合と何ら変わりなく、コミュニティ全体からの評価に耐えられるように書き上げなければならないという意識（常識）の下でコンテンツ（論文原稿、データ等）の質が保たれており、そのことこそが「研究者の作法であり暗黙の約束」である。その上で、質の保証に関しては、査読付きジャーナル論文と比較してプレプリントに固有の（あまり語られない）観点もあることを指摘しておく。査読付きジャーナル論文の場合、たとえ査読に一定程度の時間を要するにしても、それが適切になされるのであれば、質の良くない論文原稿がそのままの形で世に出ていく事態を防ぐことができる。これに対して、プレプリントの場合は必ずしも適切なチェックが入っているとは言えない。そのことの直接的な帰結は、査読付きのジャーナル論文に比べてプレプリントの質が劣っているという確定的な結論ではなく、あくまで質の評価が‘放置’される可能性が高いということである。そしてその放置されたプレプリントの質評価の結果は、現実には玉石混淆であり得る。つまり、‘玉’と評価されるべきプレプリントが見出されずに（あるいはその‘原石論文’が磨かれないままとなり）適正な評価や注目を与えられないことによる機会逸失もあり得れば、‘石’と評価されるべきプレプリントが大量に投稿されること（さらには「悪貨が良貨を駆逐」してしまう可能性）もあり得る。後の図 7 にも見るとおり、プレプリント利用の急速な拡大に伴い、被引用

機会も急拡大している中、本来のサイエンス・メリットに見合わない被引用回数を稼ぐケースも多くなり、見かけの質評価の結果に大きなバイアスを生むことがあるほか、同一の分野であっても時代が少し違えば被引用回数自体がそもそも適切な評価指標とならないことも考えられる。今後膨大な量の（場合によっては‘石’以下の）プレプリントが生産され、流通していくであろう中、ジャーナル側としてもそれを査読プロセスに載せていく場合にどのように適切にスクリーニングしていくか、そのために適切なレビューをどう確保していくかという課題は一層難しさを増していくだろう。同時に、研究コミュニティの側でも、そうした加速度的に氾濫する文献の中から価値あるものを適切に選び取っていくことが求められるという難しい時代（cf. ‘Preprint War’ [31]）に向かっていくことは想像に難くない。

1.3 プレプリントの利点・有用性

プレプリントの利点としては、研究者・執筆者にとっては何より、研究成果の迅速な発信と幅広い共有、そして研究成果についての先取権の確立を可能ならしめる点が挙げられる。査読に要する時間幅（概ね数ヶ月～数年）は、研究分野にも依るものの、各分野における研究進展との相対的な時間スケールにおいて様々な機会逸失の可能性をほらむほどに長いこともある。その結果、研究成果の先取権・クレジットが守られなくなる可能性があることに加え、限られた時間内での十分な研究業績の公開・アピールがかなわないことが（特に若手の）研究者のキャリアパス上の障害要因ともなり得る。また、査読時間の長さは別としても、研究者にとっては自らの成果がジャーナル側の査読プロセスのうちに何らかの形で‘スクープ’されることが潜在的な懸念材料として常に存在する。こうした状況を打破する手段を与えてくれることが、研究者がプレプリントを活用することの最も大きなインセンティブとなっていると言える。数ヶ月から数年を要し得る査読・出版プロセスを待つ必要がなくなり、透明性や公平性が確保された中で研究成果の発信・共有の大幅なスピードアップが可能となることは、特に厳しい時間的制約にさらされている学生やポストドクター等の若手研究者にとってはタイムリーな業績や活動状況の発信を通じて貴重なジョブ・アピールの機会ともなるほか、タイムリーな共同研究者の獲得につながることもあり、若手研究者のキャリア形成上も一層重要になっている（cf. [9]）。また、先にも触れたとおり、最新の研究動向（ライバルの動向も含め）の把握手段ともなるほか、所属する研究機関やファンディング・エージェンシー（FA）等向けの各種申請・実績報告プロセスにおいては、各機関のポリシーに応じて研究者の活動実績・活動状況を報告する上でのエビデンスとしても活用され得る⁴。また、ジャーナル誌への投稿に先立ってプレプリントの形で公開することで、研究コミュニティから迅速かつ幅広いフィードバックを受けることが可能となり、その結果ジャーナル誌への投稿時までに原稿をより質の高いものへと改善できる機会となることも利点として挙げられる。

以上は研究を発表する研究者個人にとってのプレプリントの利点であるが、研究コミュニティにとっての利点は言うまでもなく、オープンサイエンス時代にあって研究分野の進展を大きく加速させ得る点にある。個別のジャーナル側で出版情報を逐一チェックする必要がなく、基本的な情報収集が全てプレプリントサーバー側のワンストップで済むとなれば様々な研究活動の効率が大幅に上がる。加えて、コミュニティ全体にとっての研究プロセスの透明性や公平性の確保の観点もある。多くのプレプリントサーバーではプレプリントのバージョン管理がなされており、一度投稿されたプレプリントも、その後コミュニティ内の研究者から寄せられたコメントや研究の進展状況、引用文献の追加の必要性等を踏まえて随時更新・再投稿することができ、その都度プレプリントのバージョンが上がっていく仕組みになっている。その際、古いバージョンのプレプリントも削除されることなく掲載され続けることが普通であり、著者が自らの意向とタイミングでプレプリントを随時更新していく過程そのものがオープンな履歴として残り続ける。プレプリントの持つこれらの特徴は、透明性や公平性の観点から、通常の査読付きジャーナルを通じた出版プロセスにはない利点と言える。さらに、研究機関側や学協会・研究コミュニティ側、政策関係者・IR（Institutional Research）関係者にとってのプレプリントの利点・有用性も様々である。オープンアクセス（閲覧に際して無料）であることの利便性・経済性（購読契約の場合と比べて）に加え、ジャーナル・ベースの場合よりもタイムラグの影響を抑えた形で、国毎や大学・研究機関毎のパフォーマンスや分野毎の活動状況、研究動向、研究者毎の活動履歴（分

⁴ 研究評価におけるプレプリント情報の活用可能性については本稿最終節の小節 3.1 でも再考する。

野, 研究インパクト, 所属, 共著関係等), そして研究人材の把握・評価を行うことができる。

また, 前小節 1.2 で触れたプレプリントの「質」についても, 「ジャーナル側で査読がなされていないから質が低い」との指摘は必ずしも当たらないどころか, 場合によってはむしろ逆である可能性も指摘できる。それは, 出版前にプレプリントの形態を通して多くの読者(研究者)の目に触れ(いわば研究コミュニティ全体が査読者), 科学的な誤謬からアカデミック・インテグリティに至るまで様々な観点からチェックやフィードバックが入り, 研究に評価や批評を加えることが可能だからである。これは, ごく限られた数の担当編集者や担当査読者による必ずしも多面的でも網羅的でもない, どうしてもバイアスの掛かりがちな査読を通過して掲載に至ったジャーナル論文よりも, 場合によってはプレプリントのほうがより多くの批判に耐えて磨かれたステータスであることもあるという見方である。実際, 計算機科学や数学等の分野では, 一定の研究成果が出たら, 研究者はそれをまずプレプリントで晒してみても, その価値の確認・検証はそのプレプリントを通じて研究コミュニティ全体からなされるという形態が比較的定着しているという。同じく前小節 1.2 の終わりで, 研究コミュニティにおいて適切な文献を選び取っていかねばならないことの難しさについて触れたが, 研究コミュニティ自体がそうした「適切に選び取る」経験値を既に十分に獲得できている場合には, プレプリントという仕組みの存在が, 出版社におけるより質の高いジャーナル論文の輩出に貢献しているとも言える。さらには, 研究分野によっては, そもそもジャーナル誌上での論文出版を必ずしも前提としない自律的な分野発展基盤の構築・維持に貢献しているとも言える(特に情報科学系分野: 本稿 1.5 節を参照)。

加えて, 多少異なる観点としては, 通常のジャーナル側での査読では通りにくい可能性のある, 統計的な有意性が伴わない研究結果の報告や, 既存の通説に対してネガティブなアイデア, 新奇なアイデアなども, プレプリントであれば万人の目に触れ得る形で公開することができる。その際, ジャーナル論文に出す場合の編集部での土地勘や関心の有無, 査読者の確保のしやすさ・しにくさ等に由来するネガティブな要因を排除ないし緩和する効果も期待される。他にも, 学生が収集したデータ, まだ全体として完全なストーリーにはなっていない状態の原稿, 既存の研究成果と同等のものを再現したとの報告, 実施者側において必ずしも十分な評価や価値判断の伴わない研究成果等の中にも, 他者にとっては有用なものもあれば, 集積したり組み合わせたりすることで新たな価値につながるような断片的成果の類もあり, これらも公開されなければ無いも同然となってしまう。こうした成果やアウトプットについてもプレプリントの形で公開することで, その科学的価値の評価を広くタイムリーに問うことができるため, そのことで分野の本質的な研究進展やブレークスルーが支えられることもある。さらに, 一定期間の経過後であればジャーナル側での査読を経たもの(従ってジャーナル掲載版)と同じ論文原稿をそのままプレプリントサーバーに載せることが許容されているケースもある。その場合には, 研究者側もプレプリントサーバー上の最終論文を入手すればジャーナル論文を直接参照せずとも研究成果に触れられるため, 多くの研究分野で電子ジャーナルの価格高騰問題が顕在化している中, プレプリントサーバーや機関レポジトリに掲載された「最終版」の論文が研究活動を支えることも珍しくない⁵。実際, 最近 NISTEP の行った我が国の研究者等を対象としたアンケート調査 [32] の結果からは, 回答者の 8 割以上がプレプリントをインターネット上の検索エンジンを通じて入手していることが示されており, このことから「paywall の向こう側」にあるジャーナル論文が入手しにくい場合の代替手段をプレプリントが担っていることが伺える。

1.4 プレプリントのリスク・課題

プレプリントに関してよく指摘されるリスクやデメリットの観点として, まず一点目に挙げられることが多いのが, 「査読プロセスを経ていないために質が保証されていない・質が劣る」というものである。この点については, 本主張の当否も含め既に述べたとおりである。端的に言うならば, プレプリントのほうがジャーナル論文よりも質が劣るとの確固たるエビデンスはなく, アカデミアにおいてもそのようなコンセンサスではないこと, また, プレプリントの形で先行公開することによってむしろ質の確保・向上につながる可能性があることについて述べた。したがって本稿では, こうした「質」の観点を必ずしもプレプリント(のみ)に付随するリスクやデメリットとしては扱わない。

⁵ プレプリントを通じた研究活動がいかに盛んになり, 実質的な研究活動はプレプリント・ベースで完結し得るようになって, ジャーナル論文という形態が(少なくとも当面は)なくなることがないであろうと信じられる理由として, 「研究者の美学」を挙げた有識者もあった [23]。自分の論文が製本されて「形」になることに価値を感じる美学があるという。ただ, こうした「美学」そのものも時代によって変わっていくものかもしれない。

二点目として、アカデミアの範囲を越えた広く世の中との間の科学技術コミュニケーション、特にリスクコミュニケーションの観点が挙げられる。先のとおり、COVID-19の感染拡大抑制及び治療法確立に向けた研究論文のオープン化と公表迅速化が急加速しており、プレプリントサーバーはその動きを大きく後押ししてきた。しかしながら、まさに今回のCOVID-19関連研究がそうであったように、人の生命や健康等に直接影響し得る臨床医学系の研究分野等のプレプリントの場合は特別な注意が必要である。それは、必ずしも専門的知識やリテラシーを持たない非専門家も含め誰でもそのままの形でアクセスでき、場合によっては必ずしも科学的でない価値判断や期待を伴って急速に拡散してしまう可能性があるからである。SNS等の蔓延した情報化社会の中であって、この問題は一層難しさを増している。メディアを含め世の中の多くの読者は、プレプリントと査読付きジャーナル誌上で発表された論文との差をそれほど意識することなく、同等に「(信頼に値する)研究成果」として取り扱うことがあり得るからである。今般のコロナ禍に際しては特にそうした‘誤解’の生じることがないよう、図2 (bioRxiv)、図3 (arXiv)にも見るとおり、COVID-19研究関連のプレプリントを掲載するプレプリントサーバーの多くでは、トップページの目立つ場所に注意書き(プレプリントで報告されている内容はまだ科学的知見としての正確さや有用性等について確定的な評価のなされたものではないこと、したがってそのプレプリントの内容を診断・治療の参考としたり、確固たる科学的知見としてメディア等で取り扱ったりすることのないよう)が付されるようになってきている。また、プレプリントで公表される研究の中身が世間的に関心の高い社会課題に近い場合には、特定の立場や価値観をサポートするような偏った(アカデミアの適正なピアレビューを経ていれば却下されているはずの)内容がれっきとした研究者の論文として公に出回ることによって問題が‘政治化’することも考えられる。以上の各ケースに見るとおり、近年のプレプリントの台頭は、研究の成果や学術情報の信頼性をどのように担保していくか、誤った情報の拡散をどう防いでいくか・緩和していくかという問題を改めて強く意識させるに至っている。

三点目として、持続的な運営基盤の観点がある。プレプリントサーバーが学術情報の有用な共有プラットフォームであり続ける上では、DOIあるいはそれに準じる独自の文献IDが振られた上で恒常的に安定した運営がなされていくこと、そしてプレプリントが引用可能な文献としてそのIDとともに永久に保証されていくことが必要条件である。海外の大手出版社の運営するコマーシャル・ベースのプレプリントサーバーは別として、アカデミア・ベースのボランティアな運営によるプレプリントサーバーの場合は、その財政基盤の弱さが持続的運営にとって大きな課題となっている。最大かつ最古のプレプリントサーバーであるarXivも、現在では複数の収入源(米コーネル大学図書館、米サイモンズ財団、そして各国のメンバー機関⁶)からなるファンディング・モデルの下に運営されているが、その規模が進展・拡大していくにつれて、ホスト機関(2001年までは米ロスアラモス国立研究所、それ以降は米コーネル大学図書館)は持続性の伴う形での運営資金に腐心してきた。国・地域特化型の一部の新興サーバーを含め、資金面での自立的・持続的運営の難しさから経営難に陥っているプレプリントサーバーの状況も報告されている[33]。また、大手学術出版社の提供するサービスとプレプリントサーバーとの棲み分けや共存関係の在り方も単純ではない。オープンサイエンスの不可逆な潮流にあって、出版社側もプレプリントとうまく付き合うよう経営戦略を変えてきた。人文・社会科学系分野を主な対象として発展したプレプリントサーバーであるSSRN[34]を2016年にエルゼビア社が買収して自社サービスの一環に取り込んだことは記憶に新しい。他方、研究者側では、プレプリントサーバーが出版社という営利団体側に入ってしまうことで、arXivのようなコミュニティベースで運営されている場合と比べて自律性や透明性が失われてしまうことを憂慮する声も聴かれる。学術出版社側でも研究者側でも、様々な形態のプレプリントサーバーとうまく付き合っていく・使いこなしていくことが求められる時代を迎えていると言えるだろう。

四点目として、一度プレプリントの形で公開してしまうことが、その後の査読付きジャーナル誌への投稿・掲載可能性を制限してしまう(ジャーナル側から拒否される)可能性のあることが挙げられる。ただし、加速するオープンサイエンスの潮流にあって、現在では多くのジャーナル誌がプレプリントを経ての論文投稿を受け入れる、あるいはむしろ推奨するようにポリシー[11]を変更・設定するようになってきたことから、以前ほどは主たるリスクとしては語られなくなった印象がある。一部のジャーナルではいままもプレプリントで先行公開された論文を受け付けない(制

⁶ 本稿執筆時点の2020年10月現在、日本からはNIIコンソーシアムの下、計16機関(京都大、東京大、名古屋大、大阪大、東北大、KEK、北海道大、慶応大、九州大、国立天文台、早稲田大、広島大、神戸大、東工大、東京理科大、筑波大; arXivウェブサイト(<https://arxiv.org/about/ourmembers>)における掲載順)が参画している。

限する) ポリシーを採用しているが, 研究者側がプレプリントを利用した研究様式の数々のメリットに鑑みて, それと適合性・両立性のあるジャーナル誌を選ぶようになれば, 出版社側もそれに対応せざるを得ず, いずれにしろ研究コミュニティ全体としては pro-プレプリントに移行していくと考えられる⁷.

1.5 情報科学系分野の特性

本節の最後に, 特に情報科学系分野に特有の話題や観点について整理しておくことで, 次節で見る定量的分析の結果の解釈等に当たっての基本情報ないし前提の背景とする. なお, ここでの内容は主に有識者ヒアリング [24] で聴かれた指摘等に基づく.

(1) 歴史的経緯

情報科学分野では, プレプリントはごく最近, 10 年ほど前から機械学習の研究が盛んになったことで急速に利用されるようになったが, より古い時代, インターネットの普及以前は「テクニカルレポート」が今までいうプレプリントの役割を果たしていた. まだ評価の確立していないものはテクニカルレポートとして出すといった独自の文化があり, 当時は研究会や会議に出したものがそのままテクニカルレポートになったりした^a.

^a 有名どころでは, Google の創設者であるラリー・ページ氏らが開発し, Google の検索エンジンの基礎となったと言われる PageRank についての論文 [35] も初めは米スタンフォード大のテクニカルレポートとして出されている.

(2) プレプリント活用のインセンティブ

情報科学分野でのプレプリント利用はここ 10 年で急激に進んだ. それはあまりに機械学習分野の進展が早く, 年に数回程度開催される国際カンファレンスを通じた発表の場では圧倒的に足りなくなってきたからだ. そこでも発表者や論文についてはピアレビューによる厳しいスクリーニングがなされ, 採択率は 10% 以下程度という. それで, 新しい着想や結果があればまずは arXiv にプレプリントで出すという成果の共有様式が定着してきた. 査読付きジャーナルやカンファレンスの場合, 実際には価値ある論文であっても採択率が低ければ日の目を見ないため, 減点要素のない論文でないとは通らないのに対し, プレプリントであれば多少の粗があっても面白ければ価値が見出され得るため, まずは出してみようという話になる. 先に arXiv に出したことが, その後のカンファレンスでの発表を妨げるようなこともない.

(3) 研究業界の風土

近年のディープラーニング系の研究では, 何か面白そうな報告がプレプリントで出ると, 研究者やユーザーがすぐに手元で実装して, その価値を実際に試してみることができる. コードもプレプリントの中に掲載されているので, すぐにそれが実行できる. つまり, 他分野の査読付きジャーナルでいうところの査読というプロセスが要らないといえれば要らない. 自分でチェックして試せて評価できるので. 加えて, 情報系ではいわゆる「ジャンク論文」に対しても比較的大らか. 新しいコード (アイデア) というものは「当たるも八卦当たらぬも八卦」であり, 一部の他分野とは違い, 当たらなくても世の中に実害をもたらすようなシリアスな事態にはならない. たとえ間違えて遅いアルゴリズムを報告しても, そのことで社会的な迷惑にまでは及ばない. それで, とりあえずプレプリントで出しておこうという話になる.

⁷ プレプリントについて最も導入や適応の早かった高エネルギー物理学分野であっても最初期には辿らざるを得なかった道と聴く. オープンアクセスを志向する研究者側と旧来の論文出版ビジネスを擁する大手学術出版社側との間のテンションに関しては, シュプリンガー・ネイチャー社の新雑誌 (Nature Machine Intelligence) への関与 (投稿も査読も) を AI 関連の研究者らがボイコットする輪が広がった事例 (2018 年) も記憶に新しい (<https://openaccess.engineering.oregonstate.edu/home>).

(4) ジャーナル論文との関係

情報系の分野には（そこに載れば高く評価されるような）圧倒的に有力なジャーナルというものがない。*Nature* や *Science* にも情報系の論文が載ることはあるが、そこに載ったからといって評価されるというものでもない。やはりカンファレンスでの実績というものが重要になる。ジャーナル論文であるかどうかはあまり関係なく、プレプリントであってもカンファレンス論文で引用されれば正当に評価される。他方で、（アカデミアのジョブマーケットで競うわけではない）*Google* の出す論文などはカンファレンスに出す動機もなく、プレプリントに出してお仕舞いとなることもある。

以上、端的には、他者の研究成果について研究者自身で実際に手軽に検証・評価できること、そして研究や開発のサイクルが他の研究分野に比べても特に速いことが、情報系分野に独自のプレプリント利用文化が根付いてきたことの背景にあると言える。この様子は実際の arXiv のデータをもとにした次節の定量的分析においても傾向として浮かび上がるとおりである。

2 知の生産活動におけるプレプリントの役割（定量的分析）

プレプリントをめぐる動向は、アカデミアと社会とのより良い共創関係を目指す科学技術・イノベーション政策上も有用な情報源となり得る。例えば、研究評価に当たっては、いわゆる Top 10% 論文（各論文について）やインパクトファクター（各ジャーナルについて）等の定量的指標がこれまで広く使われてきたが、そこではプレプリントの関与する各種の活動状況は反映されてこなかった。ジャーナル論文と合わせてプレプリントの情報までをいかに考慮していくべきかについて今後検討を進めていくに当たっては、以下の問：

プレプリントを通じた研究様式は、現代の研究活動（論文引用による新たな知の生産活動）において、実際にどれほどの存在感・インパクトを担っているのか

に対する答について、定量的な裏付けを持って把握しておくことが EBPM (Evidence-Based Policy Making) の前提として欠かせない。ただし、この問に対する具体的な答は現時点において（世界的にも）見当たらず、プレプリントサーバーに眠る情報を有用な情報源として活かし切れていないのが現状である。本節では、この間に向き合い一定の回答を与えるべく、まずは最も歴史があり投稿数も多いプレプリントサーバーである arXiv [6] (図 3) を対象に、そのデータを活用して計量書誌学的アプローチにより試行的に分析した結果について報告する。

2.1 分析手法

被引用回数ベースの定量的指標は、特に自然科学系の分野において研究論文の「質」を測る上での有用な媒介変数としてしばしば用いられ、知の生産活動における一種のインパクト指標と見なされてきた [36]。本稿では、ジャーナル論文に留まらずプレプリントまでを含んだ形での被引用状況の分析を通じて、現代の知の生産活動の全体像及びその中でプレプリントの担っているインパクトを一部なりとも定量的に描き出すことを目指す。その際、被引用回数ベースの新たな指標を提案し、これをもとに（数物系の枠内ではありつつも）分野横断的な議論を試みる。

先の表 1 に見るとおり、プレプリントサーバーによっては、プレプリントに DOI を付与しているものもあれば、そうでない独自の ID によりプレプリントを管理しているものもある。arXiv は後者に該当し、まだジャーナル論文等として出版されていない場合には DOI 情報が付加されない⁸。そこで、あるプレプリントに着目したとき、それがプレ

⁸ 今後、arXiv のプレプリントにも DOI が付与されていく可能性もあるという（arXiv 運営側で議論が継続されている状況）。

arXiv is a free distribution service and an open-access archive for 1,771,036 scholarly articles in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics. Materials on this site are not peer-reviewed by arXiv.

Subject search and browse:

News
 arXiv now processes new submissions and replacements with TeX Live 2020. [Learn more.](#)
 Read about recent news and updates on [arXiv's blog](#). (View the former "what's new" pages [here](#)). Read [robots beware](#) before attempting any automated download.

COVID-19 Quick Links
 See COVID-19 SARS-CoV-2 preprints from
 • [arXiv](#)
 • [medRxiv and bioRxiv](#)
Important: e-prints posted on arXiv are not peer-reviewed by arXiv; they should not be relied upon without context to guide clinical practice or health-related behavior and should not be reported in news media as established information without consulting multiple experts in the field.

Physics

- **Astrophysics (astro-ph new, recent, search)**
 includes: Astrophysics of Galaxies; Cosmology and Nongalactic Astrophysics; Earth and Planetary Astrophysics; High Energy Astrophysical Phenomena; Instrumentation and Methods for Astrophysics; Solar and Stellar Astrophysics
- **Condensed Matter (cond-mat new, recent, search)**
 includes: Disordered Systems and Neural Networks; Materials Science; Mesoscale and Nanoscale Physics; Other Condensed Matter; Quantum Gases; Soft Condensed Matter; Statistical Mechanics; Strongly Correlated Electrons; Superconductivity
- **General Relativity and Quantum Cosmology (gr-qc new, recent, search)**
- **High Energy Physics - Experiment (hep-ex new, recent, search)**
- **High Energy Physics - Lattice (hep-lat new, recent, search)**
- **High Energy Physics - Phenomenology (hep-ph new, recent, search)**
- **High Energy Physics - Theory (hep-th new, recent, search)**
- **Mathematical Physics (math-ph new, recent, search)**
- **Nonlinear Sciences (nlin new, recent, search)**
 includes: Adaptation and Self-Organizing Systems; Cellular Automata and Lattice Gases; Chaotic Dynamics; Exactly Solvable and Integrable Systems; Pattern Formation and Solitons
- **Nuclear Experiment (nucl-ex new, recent, search)**
- **Nuclear Theory (nucl-th new, recent, search)**
- **Physics (physics new, recent, search)**
 includes: Accelerator Physics; Applied Physics; Atmospheric and Oceanic Physics; Atomic and Molecular Clusters; Atomic Physics; Biological Physics; Chemical Physics; Classical Physics; Computational Physics; Data Analysis, Statistics and Probability; Fluid Dynamics; General Physics; Geophysics; History and Philosophy of Physics; Instrumentation and Detectors; Medical Physics; Optics; Physics and Society; Physics Education; Plasma Physics; Popular Physics; Space Physics
- **Quantum Physics (quant-ph new, recent, search)**

Mathematics

- **Mathematics (math new, recent, search)**
 includes (see detailed description): Algebraic Geometry; Algebraic Topology; Analysis of PDEs; Category Theory; Classical Analysis and ODEs; Combinatorics; Commutative Algebra; Complex Variables; Differential Geometry; Dynamical Systems; Functional Analysis; General Mathematics; General Topology; Geometric Topology; Group Theory; History and Overview; Information Theory; K-Theory and Homology; Logic; Mathematical Physics; Metric Geometry; Number Theory; Numerical Analysis; Operator Algebras; Optimization and Control; Probability; Quantum Algebra; Representation Theory; Rings and Algebras; Spectral Theory; Statistics Theory; Symplectic Geometry

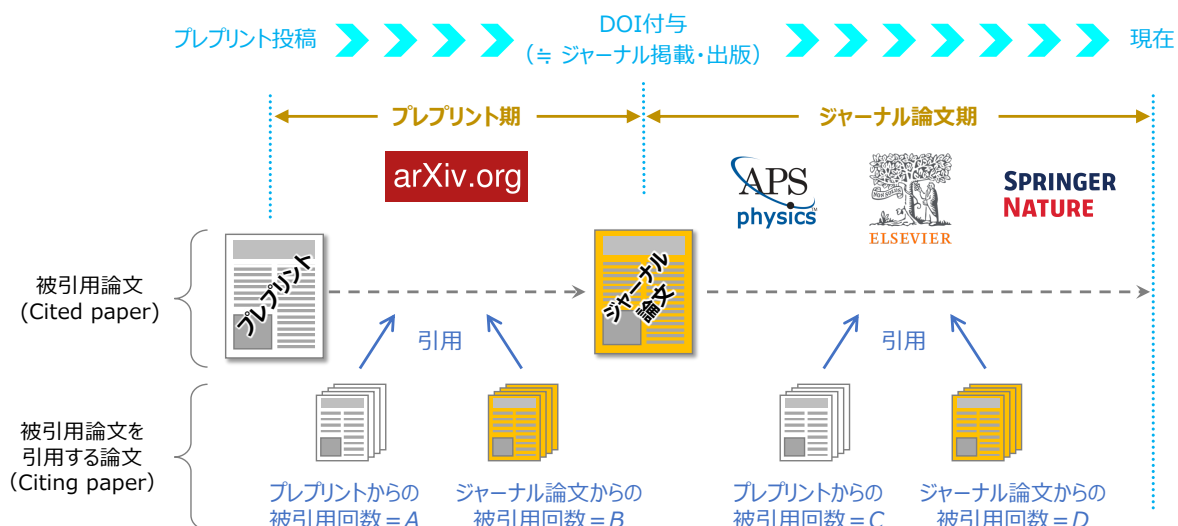
【図 3】 arXiv ウェブサイト (トップページ) [6]

プリントサーバーに投稿されてから何らかの媒体で出版されるまで、つまり DOI が付与されるまでの期間と、DOI が付与されてから現在 (データ取得時点: 2020 年初め) に至るまでの期間とを、それぞれ「プレプリント期」⁹、「ジャーナル論文期」と呼称することにする¹⁰。また、同じくあるプレプリントに着目したとき、それがデータ取得時点において獲得している総被引用回数のうち、それがプレプリント期にある文献からの引用なのか、あるいはジャーナル論文期にある文献からの引用なのかを区別して分析する。図 4 にその概念図を示す。図中の被引用回数の説明において、 A と B のカウント、そして C と D のカウントを区別している「プレプリント」と「ジャーナル論文」の区別は、データ収集時点でのステータスに基づくものである。その上で、本稿の関心対象とする新指標は以下の三つである:

指標 1	総被引用回数に占めるプレプリント期に獲得した被引用回数の割合: $\alpha (\%) = \frac{A+B}{A+B+C+D}$
指標 2	総被引用回数に占めるプレプリントから獲得した被引用回数の割合: $\beta (\%) = \frac{A+C}{A+B+C+D}$
指標 3	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: $\gamma (\%) = \frac{A+B+C}{A+B+C+D}$

⁹ 「プレプリント期」にある論文の中には、査読付きジャーナルへの投稿準備中のもの、既に投稿済みで査読プロセス中のもの、あるいは既に受理された後で出版準備中のものもあれば、査読付きジャーナルへの投稿・掲載を元より念頭に置かず arXiv を一種のリポジトリとして利用しているだけの電子プリントの類も含まれる。最後のケースを「プレ」プリントと呼称することは厳密には当たらないものの、本節における分析では、DOI の付与されていない論文については一律に「プレプリント」と呼び、そのステータスにある時期を「プレプリント期」と呼称している。

¹⁰ ジャーナル論文以外にも DOI の付与された文献も存在するほか、ジャーナル論文として出版されていても DOI 付与のない文献も存在し得るが、ここでは便宜上、DOI 付与をもってジャーナル論文としての出版ステータスの代理変数とする。



【図4】 被引用回数のカウントに関する整理

指標 1 は、論文が被引用を獲得するにあたってプレプリント期がいかに重要な稼ぎ時であるかを表すものと言える。また、指標 2 は、論文にとってプレプリントがいかに重要な被引用の獲得源であるかを表すものと言える。そして指標 3 にはこれらの両観点が含まれており、最も広い意味でプレプリントのインパクトを捕捉できる指標となっている。これらの指標設定のもとで分析を行った結果、指標 1 の値が $\alpha\%$ 、指標 2 の値が $\beta\%$ 、指標 3 の値が $\gamma\%$ と算出されたならば平均的には以下のとおり結論付ける（推定する）ことができる：

1. 文献の獲得する被引用回数のうち $\alpha\%$ は未出版の時期に獲得している。
2. 文献の獲得する被引用回数のうち $\beta\%$ は未出版の文献から獲得している。
3. もしプレプリントという様式が存在しなければ（つまり、ジャーナル論文がジャーナル論文を引用するという形でしか文献引用がなされないのであれば）、潜在的に獲得可能であった被引用回数のうち $\gamma\%$ 分を失うことになる。言い換えれば、文献引用を通じた知の生産活動の $\gamma\%$ はプレプリントなしには成立していない。

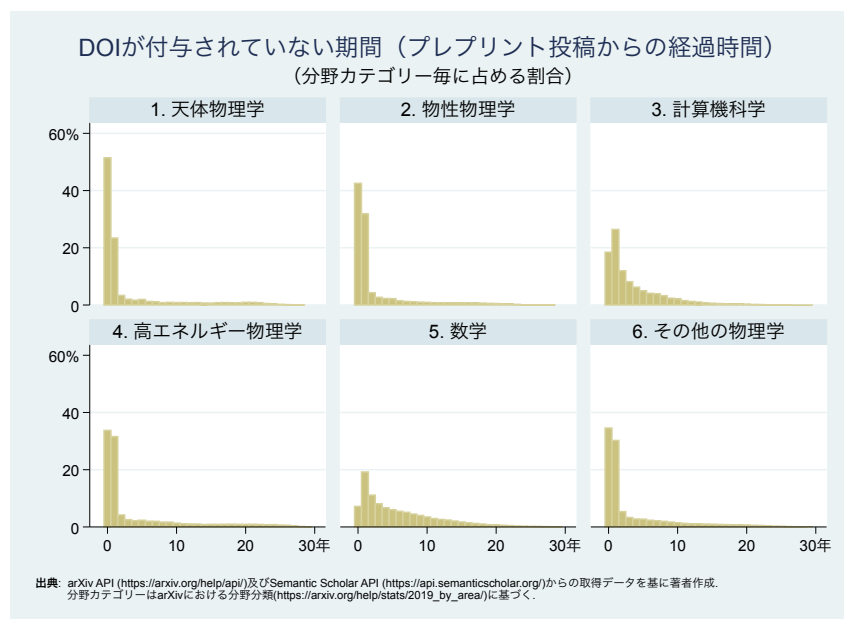
本稿第 2.3 節では実際の arXiv データについて α - γ 値を算出した結果をもとに議論を行う。

2.2 データ

本稿における分析で使用するデータセットは文献 [37] で使用されたものと同じであり、データの取得法等についての詳細はそちらを参照されたい。以下では本節での分析結果を解釈する際に必要となる基本的な事項についてまとめておく。

arXiv データとしては 2020 年 1 月 21 日時点で収集可能なものを全収集しており、2020 年 1 月 17 日までに投稿された計 1,622,763 件のプレプリント情報を使用している。被引用回数についてのデータは 2020 年 1 月 24 日から 2 月 7 日までの期間で Semantic Scholar API を通じて取得している。いずれも本稿における分析に当たっては「年」を時間に関する最小粒度として扱う。ここで、プレプリントの投稿年と DOI 付与年とが同じ場合、当該年に獲得した被引用回数は、本稿ではプレプリント期に獲得されたものと整理して分析を行う¹¹。被引用回数は Semantic Scholar 側で同定されており、プレプリントがその後ジャーナル論文に採録され、その形で引用された場合でも同じプレプリント

¹¹これをジャーナル論文期に獲得されたものと整理して分析を行った場合にも定性的な結論は本稿で得られるものと変わらないことを付記しておく。



【図5】 「プレプリント期」の長さ

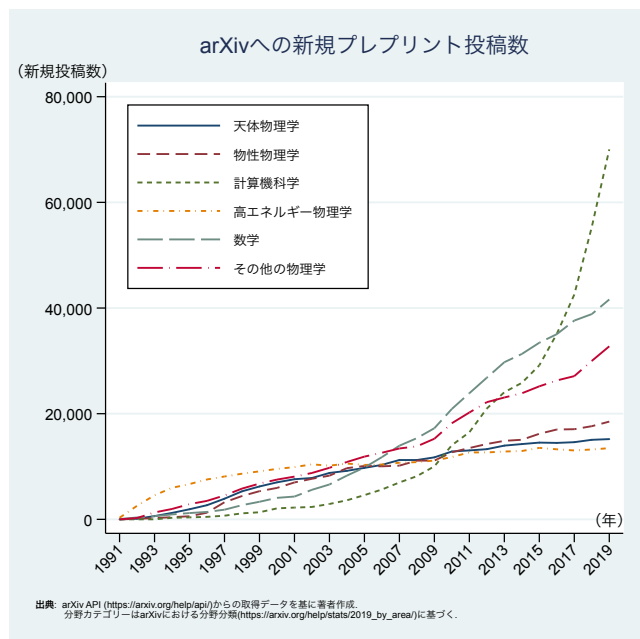
IDにおいて一括管理されている。分野カテゴリーの分類としては、arXivで使用されている153分野分類に基づき、arXivの統計情報サイト [38] で利用されているものと同じ粒度で表2のとおり6分野に大括り化したものを採用した。その際、arXivでは一つのプレプリントに複数の分野カテゴリーを設定（cross-listing）できるため、複数の分野カテゴリー属性を持つプレプリントは個々の分野の一つずつカウントした。

【表2】 6分野カテゴリー（arXiv分類の大括り化）

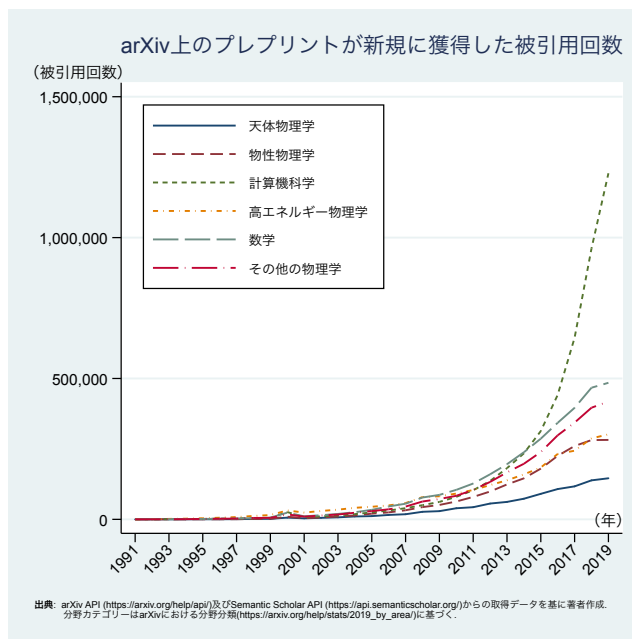
分類カテゴリー	arXiv上の分類
天体物理学	Astrophysics (astro-ph*)
物性物理学	Condensed Matter Physics (cond-mat*)
計算機科学	Computer Science (cs*)
高エネルギー物理学	High Energy Physics (hep-th + hep-ph + hep-lat)
数学	Mathematics (math + math-ph)
その他の物理学	physics + nucl + gr-qc + quant-ph + nlin

2.3 分析結果及び考察

図5はプレプリント期（DOIが付与されていない期間）の長さの分布を分野毎に表したものであるが、分野間で顕著な傾向差があることが見て取れる。物理学カテゴリーの分野群では、プレプリント投稿から概ね2年間ほどで過半数のプレプリントに対してDOIが付与されているのに対し、計算機科学や数学では投稿直後の数年間に対する集中度合いやその後の減衰の度合いが緩やかである。計算機科学や数学ではそもそもDOIが付されないままに現在に至るプレプリントの割合が大きい（計算機科学：約69%、数学：約77%）[37]。そのようなプレプリントが毎年積み上がっていくために、図5に見るとおり、これらの分野については経過年数に対してheavy-tailedなヒストグラムとなっている。他の分野が概ね半年でDOI付与に至っている中、数学分野では概ね1年程度がかかっていることが文献[37]でも指摘されている。以降の分析や考察において、被引用回数にまつわる各種指標について分野間比較を見ていくに当たっては、図5に見られる時間スケールの違いを念頭に置いて解釈していくことが重要となる。



【図 6】 arXiv への新規プレプリント投稿数



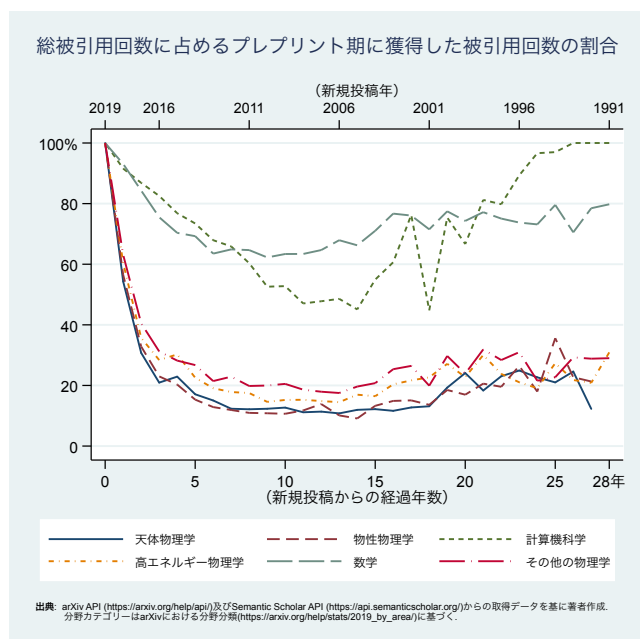
【図 7】 arXiv 上の論文が新規に獲得した被引用回数

まず arXiv への新規プレプリント投稿数の分野毎の経年推移を図 6 に示す。arXiv の statistics サイト [38] や文献 [37] でも指摘されているとおり、近年の AI 研究進展を反映して計算機科学分野カテゴリーの新規投稿数が急増している。2000 年代後半までは 6 分野の中でも最低位にあったものが、2016 年以降は他分野を大きく引き離す勢いで伸びている様子が見て取れ、2019 年の新規投稿数は約 7 万報を記録している。数学分野も 4 万報強の新規投稿数をマークしており、高度情報化社会にあつての近年の数学研究の盛り上がりを象徴していると言える。知の生産活動が活発になされている様子は、投稿論文数に加え、それらの論文が引用する文献（プレプリントも含む。）の数、すなわち分野カテゴリー毎に見たときの被引用回数からも観測される。図 7 は arXiv 上の論文が新規に獲得した被引用回数の経年推移を分野毎に表したものである。ここでの被引用回数は、プレプリント期に稼いだものに加え、ジャーナル論文として出版された後に稼いだものも含まれている。直近の 2019 年では、計算機科学分野のプレプリントや論文に対して年間を通じて計 120 万回を超える被引用がなされており、同様に増加傾向にある他分野をさらに大きく引き離す勢いで伸びている様子が見て取れる。

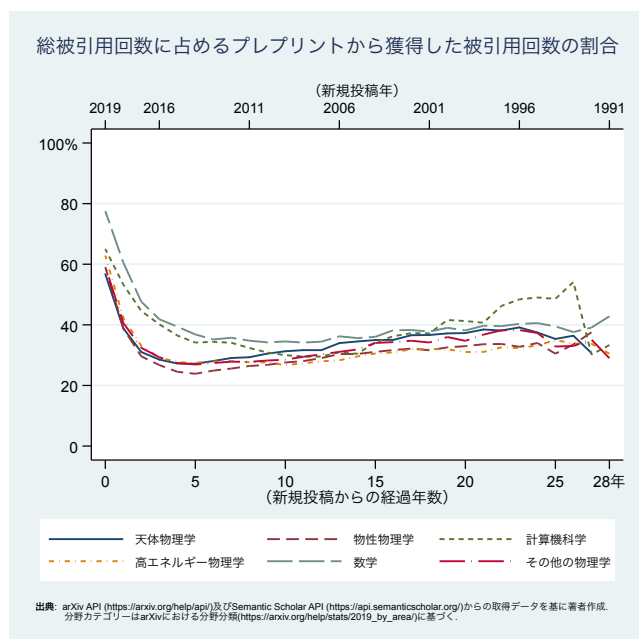
図 8 は総被引用回数に占めるプレプリント期に獲得した被引用回数の割合（指標 $1:\alpha$ 値）の経年推移を分野毎に表したものである。この指標は先のとおり、端的に言うなら、論文が被引用を獲得するにあたってプレプリント期がいかに重要な‘稼ぎ時’であるかを表すものである。例えば、論文の被引用価値が最も高い時期がプレプリント発表後のごく短い期間であった場合、それがジャーナル論文として出版された頃にはあまりもう引用されなくなっているケースもある。その場合、図 8 縦軸の値は、プレプリント投稿後の経過年数が上がっても（右に行っても）高い位置で留まり続けることになる。他方で、プレプリント期と比較してジャーナル論文期のほうが顕著に高頻度で引用される場合には、図 8 縦軸の値は低く、被引用価値のあるジャーナル論文時期が持続するほど、グラフは右下がりになる（プレプリント期の被引用回数は固定された上でジャーナル論文時期の被引用回数は増えていくため）。

まず全体に着目すれば、計算機科学・数学とそれ以外とで傾向が大きく異なる¹²。計算機科学・数学の場合は α 値が高く、特に数学分野の場合には期間を通じて平均的に 60% から 70% ほどの被引用価値がプレプリント期に集中し

¹²用語の定義上、データ取得時点において DOI が付与されていないプレプリントについては全期間をプレプリント期として扱っていること、また、分野毎に DOI 付与率は大きく異なり、特に計算機科学分野は DOI 付与率が低いとの指摘 [37] もあることから、解釈に当たってはこれらの点に注意を要する。



【図8】 総被引用回数に占めるプレプリント期に獲得した被引用回数の割合（‘ α 値’）

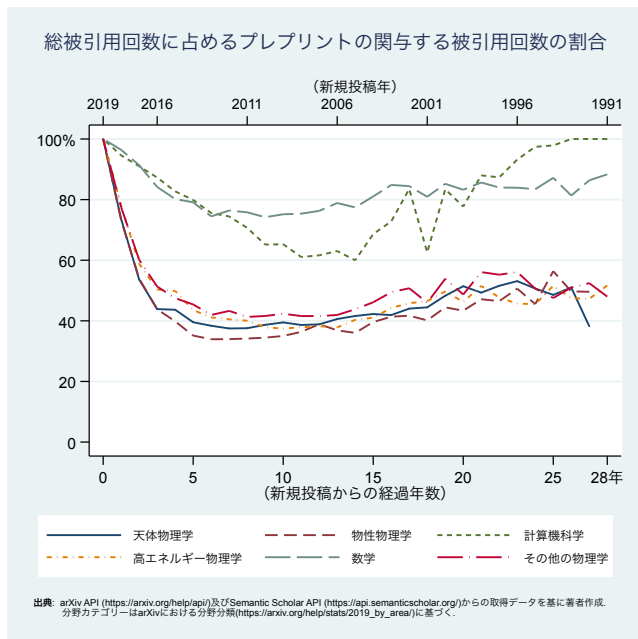


【図9】 総被引用回数に占めるプレプリントから獲得した被引用回数の割合（‘ β 値’）

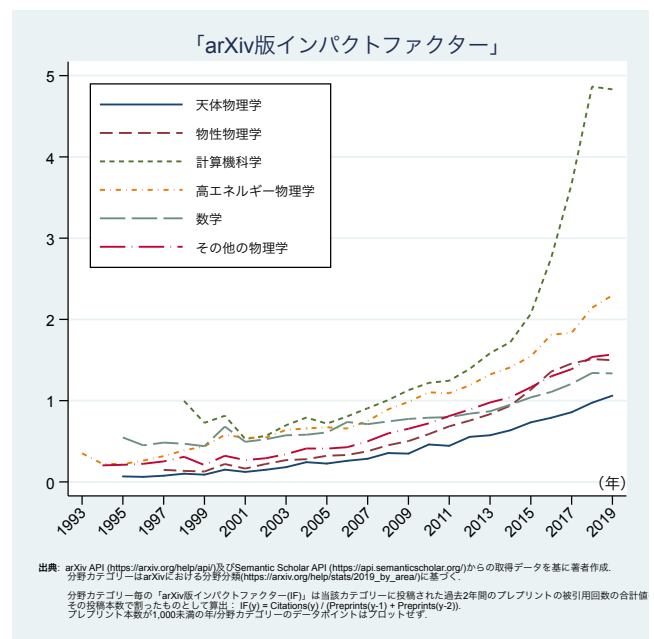
ていることがわかる。計算機科学分野についても高い値を維持しているが、2006年頃（ちょうど第3次AIブーム：ディープラーニング時代が始まった時期）を境に傾向に差が見られる。2006年以降に投稿されたプレプリント（図の左半分）は、平均的に見たとき、時間の経過に比例して α 値が下がる傾向にあり、このことは計算機科学分野のプレプリントがジャーナル論文として出版された後、10数年ほどの長期にわたって引用され続ける傾向にあることを示している。これに対し、2007年以前に投稿された同分野のプレプリント（図の右半分）は全体的に右上がりとなっており、特に1990年台（arXiv立上げ後はじめの10年）では α 値は100%に近い。他方で、物理学系の4分野については、一定程度の年数（5年程度）が経過した後は、 α 値は概ね15%~30%で推移しており、経過年数の大きな（古い）論文ほどその値は上昇する傾向にある。分野によって多少の差はあれ、いずれの分野でも α 値が一定の高さを持って推移している様子は、それだけ被引用回数を稼ぐ上でプレプリント期が重要な役割を担っているということ、言い換えれば、新たな知を生み出していく上でプレプリントが本質的に重要な役割を果たしていることの表れと言える。また、いずれの分野でもグラフが全体傾向として右下がりとなっていないことから、どの分野についても論文には被引用価値に関して‘賞味期限’が平均的に存在すること、言い換えれば、十分に長い期間が経過すればもはや引用されなくなることが示唆される。

図9は総被引用回数に占めるプレプリントから獲得した被引用回数の割合（指標2: β 値）の経年推移を分野毎に表したものである。これは、現在（2020年初頭）までに論文が稼いだ被引用回数のうち、現在に至るまでDOIが付与されずに残っているプレプリントからの被引用回数の割合を示したものである。面白いことに、一定期間（10年間ほど）以上経過した後は、多少の幅こそあれ、6分野カテゴリー間でそれほど相違なく、 β 値は概ね25%~40%の範囲に収まって推移している。その間、古い論文ほど β 値は上昇傾向にある。脚注9でも触れたとおり、本節における分析では、記録上DOI付与のないarXiv上の文献は全て「プレプリント」として扱っており、その際、DOIの付与されていない個々の理由までを区別した形での分析は行っていない。したがって、ここで見られた β 値の振る舞いについて、その理由にまで踏み込んだ解釈を行うことは難しいが、いずれにしろDOI付与のないステータスの文献が被引用回数の重要な供給源となっていることは確かである。

以上の結果も踏まえつつ、総被引用回数に占めるプレプリントの関与する（＝プレプリント期に獲得した、あるいは



【図 10】 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（ γ 値）



【図 11】 「arXiv 版インパクトファクター」

はプレプリントから獲得した)被引用回数の割合(指標3: γ 値)の経年推移を図 10 に示す。 γ 値はその定義上も明らかとなっており、概ね α 値(図 8)と β 値(図 9)の傾向を併せ持つものとなる。 γ 値の比較的安定する期間(プレプリント投稿からの経過時間が5年以上20年未満)について平均値をとれば、物理系の4分野ではいずれも4割前後、計算機科学分野では約7割、数学分野では約8割という結果になる。したがって、分析や解釈上の諸条件や制約には十分に留意する必要があるものの、端的かつ標語的にまとめるなら:

論文引用を通じた「知の生産活動」の顕著な割合—物理系の分野では約4割、計算機科学・数学分野では約7~8割—はプレプリントなしには成立していない

と言えることになる。以上の計量書誌学的アプローチに基づく試行的な分析・検証結果からは、プレプリントが過去30年近くにわたって果たしてきた、論文引用への関与という形での知の生産活動における寄与の一端が窺い知れる。従来のジャーナルベースでの学術情報の流通に依存しない形での科学の発展を arXiv が大きく支えてきた実態が定量的なエビデンスをもって初めて浮かび上がったと言える。

最後に、図 11 は今回の分析チームにおいて「arXiv 版インパクトファクター」と名付けたものの経年推移を分野毎に表したものである。これは arXiv の各「分野カテゴリー」の括りを個々の「分野ジャーナル(ただし査読は受けていない)」のように見立てた上で、それぞれ「インパクトファクター」に相当するものをアナロジーとして算出したものである。様々な研究評価に際してよく使用される Clarivate Analytics 社の Journal Citation Reports (JCR) データベースに基づくインパクトファクターでは、《あるジャーナルの影響度(インパクト)を数値化するにあたって、そのジャーナルが掲載した過去2年間の論文の被引用回数の合計値をその掲載本数で割ったものとして算出》するが、今回の指標は、その概念を拡張する形で、《arXiv のある分野カテゴリー¹³のインパクトを数値化するにあたって、その

¹³ここでは arXiv というマルチ分野のプレプリントサーバーを分析対象としたために、個々の分野カテゴリーをそれぞれ「ジャーナル」のように見立てたが、分野特化

カテゴリーに投稿された過去2年間のプレプリントの被引用回数の合計値をその投稿本数で割って算出している。今回の分析では Semantic Scholar データベースに基づく被引用回数を使って算出しているが、被引用回数のカウントの仕方は JCR データベースに基づくものやそれ以外のものも含め、提供者や利用者によって様々であるため、今回の「arXiv 版インパクトファクター」と既に流通しているそれ以外の「インパクトファクター」指標とを直接比較した形で論じることは必ずしも適切ではない。その上で、仮に一定程度でも比較して論じることに意味があるとした場合には次のことが言えるだろう。仮に今回の「arXiv 版インパクトファクター」が分野毎の平均的な査読付きジャーナルの「インパクトファクター」を超えていたなら、その査読付きジャーナルに掲載される論文よりも、その分野の arXiv プレプリントの集合のほうが、平均的に見るとインパクトが高い（査読を受けていないにも関わらず）との見方も成り立ち得る。その意味で、図 11 に見る「arXiv 版インパクトファクター」の振る舞いは示唆的であり、広く研究評価に携わる関係者の関心と呼ぶものともなるだろう。実際には、先のとおり被引用回数の依拠するデータベースの違いや、そもそもの分野カテゴリー毎のカルチャーの違い（例えば情報科学や数学の分野では査読付きジャーナルに出版する傾向が他分野と比べて低い）もあるため、丁寧な検証が伴わなければミスリーディングな議論になってしまう点に十分な留意が必要であるが、これを差し引いたとしても、これまでの「インパクトファクター」という指標への受け止めや評価に対して一石を投じ得る分析結果と言えるだろう¹⁴。

3 今後の科学技術行政への示唆

最後に、前節までに得られた定性的・定量的エビデンスを踏まえ、有識者の方々から聞かれた声も交えつつ、今後の科学技術行政への示唆について短く論じる。

3.1 研究評価の観点

まず研究評価の観点から今後の科学技術政策への示唆を論じたい。これまでの研究評価では、各種の政府目標等にもあるとおり、いわゆる Top 10% 論文やインパクトファクターといった定量的指標が幅広く使われてきた。しかしながら、これらはいずれも査読付きジャーナル誌やそれに掲載された論文を対象として定義・算出されているものであり、本稿で詳細分析の対象としたプレプリントまでを含むものではなかった。また、こうしたジャーナル論文ベースの指標をもとにした分析や評価は、計算機科学¹⁵等の一部の分野では特別なバイアスがかかりがちであった。査読付き論文のみを元に研究評価が行われる場合、それが論文の被引用回数（Top 10% 論文・Top 1% 論文であるかどうかなど）を見るのであろうと、その出版されたジャーナルのインパクトファクターを見るのであろうと、プレプリントを通じた研究活動は評価の対象から漏れてしまうことになる。例えば、論文が査読付きジャーナルに投稿され、場合によっては既に査読を通過して受理されていても、それが出版されていなければ、研究費事業の審査側にとっては直接的な評価対象とはならないこともあるのが現状である。他方で、研究活動のサイクルの速い分野であるほど、出版された内容のみを前提として今後の研究計画を立てることはもはや成立しにくく、出版論文のみを評価の対象とすることも同様に合理的でないことになる。

今回の調査・分析を通じて、プレプリントがいまや現代の知の生産活動にとって欠かせない機能を担っていること、そしてその存在感や重要性は年々増していることが、今回提案された新しい指標を通じて実際に見て取れたことは、研究評価の在り方を含む今後の科学技術行政を考える上で示唆に富む。プレプリントの役割や意義について研究者や政策関係者が半ば肌感覚で抱いてきた印象が、今回の調査によって定量的な裏付けを持って検証された意味でも意義深い。例えば、政府の科学技術基本計画においても標榜される Society 5.0 を描く上で政策的にも大きな注目と期待を

型の他のプレプリントサーバーの場合には、それを丸ごとここでいう「ジャーナル」のように見立てることで「プレプリントサーバー版インパクトファクター」を定義・算出することも考えられる。ただし、その際の留意点も本文中の arXiv 版インパクトファクターの場合と同様である。

¹⁴ arXiv データをもとにした今回の分析では、その対象とする分野特性上、バイオ・ライフサイエンス系のプレプリントのインパクトまでは分析の対象とできていないため、本節における結論はいずれもあくまで arXiv の対象とする数物系の分野に限ったものである点に重ねて留意が必要である。

¹⁵ 計算機科学（情報科学）分野では、研究評価に際して、いわゆるトップカンファレンスでの採択実績が重視され、ジャーナル論文そのものは比較的重視されないとの指摘もある [39, 40]。

集めている人工知能研究を含む計算機科学系の分野について、ジャーナル論文ベースの指標からはなかなかつかみきれない研究動向やインパクトに関する情報やエビデンスが、こうしてプレプリントのデータから新たに得られたことは、今後の科学技術政策を EBPM の観点から支えていく上でも有用である。有識者からは、

ジャーナル論文として出版されていないプレプリントであっても知の生産活動に大きく貢献していることが今回の分析結果からも示されている。これまで使われてきた各種のジャーナル論文関係の成果指標では多分に一面的な評価しかなされてこなかったのではないかと

との意見が聴かれた。あわせて、全てが客観的な指標で評価されるものでもなく一流の目利きも重要との指摘もなされた。

加えて、世界的潮流として、いまや多くの研究者が毎日更新されるプレプリントサーバーを重要な情報源としていることに鑑み、研究費申請や研究実績報告等に際しても、最新の研究動向の説明やそれをもとにした研究者個人の今後の研究計画の提案等を行う際にプレプリント情報を活用していくことが効果的である場面は今後益々増えていくと考えられる。そうすることで、タイムリーな研究動向をエビデンスに基づいて反映していくことも可能となれば、FA 等とのやりとりの中で書類上の形式や実績報告期限等を意識するあまり中途半端な出版になる、あるいは無用なコストを掛ける等の事態を回避することにつながる可能性もある。現状、世界各国の代表的な FA・研究支援団体やその研究費事業等における研究評価（申請書、中間・事後評価書等）に際してのプレプリントの扱いについては FA 毎に対処方針が異なり（cf. [8, 41]）、全体として過渡期にある状況が浮かび上がる。今後、我が国においても、プレプリントを通じた活動状況を研究審査や実績評価等の研究評価にどのように反映していくかについて、一定の考え方を整理する必要が生じてくるだろう。有識者からは、

研究評価に際してプレプリントは現状を知る情報として重要。そもそもジャーナル論文をもとにした評価に基づいて研究費がつけ始めた時点で研究テーマとしては既に‘終わっている’とも言える。そうした研究費配分では過去の後追いにしかない。研究費事業等での研究評価とは、本来的にはジャーナル論文としての出版成果を見るものではなく、申請内容が妥当であるかどうかの中身で見るべきであって、そこでプレプリントの速報性が有用なのであれば積極的に活用していくべき

との意見が聴かれた [23]。加えて、

科学の歴史とは繰り返すもの。行政はその歴史を把握して、いまはどのフェーズ、次はどのフェーズというのを知るべき。‘理論’が必要な時期に‘実験’に投資しても意味がないということ。そうした見極めに際してプレプリント情報はタイムリーかつ有用なガイドとなる可能性がある

との意見も聴かれた [23]。研究者の研究様式にプレプリントが標準的に組み込まれていること・組み込まれていくことと歩調を合わせて、あるいは先取りする形で、行政や FA 側でも今後の研究評価に際してプレプリントを標準的に組み込んでいくことを考えていくべき時機と言える。また、表 1 にも見るとおり、人文科学系や社会科学系の分野でもプレプリントが盛んに活用されている中、今後我が国においてこれらの分野新興やそれに伴う研究評価の在り方を考えていく上でも¹⁶、これらの分野特性（自然科学分野との差異）を適切に踏まえつつ、プレプリントの意義や役割について引き続き検討するとともに政策へと反映していく必要があるだろう。

¹⁶ 令和 2 年第 201 回国会において成立した「科学技術基本法等の一部を改正する法律」において、「科学技術基本法」の対象に「人文科学のみに係る科学技術」を加えるよう変更がなされている（ここでいう法文上の「人文科学」には一般の社会科学も含まれる。なお、「科学技術基本法」は「科学技術・イノベーション基本法」に変更される（令和 3 年 4 月施行）。

3.2 科学コミュニケーションの観点

冒頭でも触れたとおり、今般のコロナ禍は情報の信頼性や伝達の正確性といった観点でこれまでにない学術情報と社会との関わり方の難しさを浮き彫りにしたと言える。プレプリントが大きな役割を果たすオープンサイエンス時代にあつて、科学コミュニケーション（COVID-19 関係の研究では特にリスク・コミュニケーション）の在り方も、従来のジャーナル・ベースでの科学的知見の発信・共有に基づくものから変化していかざるを得ないだろう。特にプレプリントは、その利用者の属性や科学との関わり方（研究コミュニティの内外）をよく意識して取り扱われなければ、科学的知見に関して誤った情報を拡散させかねない危うい情報媒体・ソースとなり得ることが各所で指摘されている。実際、冒頭で触れたプレプリント撤回の事案 [25] のとおり、COVID-19 関係のプレプリントでは、必ずしも査読力や科学リテラシーの及ばない層の目にも意図されない形で触れることで、真実ではない情報があたかも真実であるかのように扱われたケースも発生し、各種の SNS 等を通じたその拡散速度も相まって社会的な問題ともなった。こうした事案に照らし、有識者からは、プレプリントの役割や意義を考える上ではその研究コミュニティの内外の境界をよく意識することが重要との指摘が聴かれた。研究分野のコミュニティ内では普通に通じることであつても、そのコミュニティの外（他の研究分野、メディア、一般の方）ではそれが通用しないことが多々あり、アカデミア内であっても研究分野が違えば）支障が生じ得る中、それがアカデミアの内と外との関係となればなおさら問題は根深いと言える。有識者からは、

一般の方々には、科学者の言うことや研究成果は正しいものだと思われている。しかし実際は、科学者の言うことには多くの仮説や誤りが含まれており、その中で検証されて磨き抜かれることで学問として発展していくもの。どのような査読を行えば正しい科学が生き残るか、どうすれば適切にふるいにかけることができるかというのは難しいこと。後に誤りと判断された論文であっても一流誌に載ることはある。正しい科学というものは科学者の試行錯誤や切磋琢磨の中で生き残っていくものであつて、何十年、何百年とかがかかって定説が形作られ、教科書にも載っていくもの。その上でさらに検証され続けていく世界。そうした科学の営みの本質についての理解がないままにプレプリントが出回ることは危うい

との意見が聴かれた [22]。あわせて、

今回の分析対象は *arXiv*（数物系）にとどまっているが、社会的なインパクトの大きさ・注目度の高さからは、やはりバイオ系・医学系のプレプリントも見えていくことが大事。数物系では見られなかった問題や難しさもはらんでいる。人の命や健康に直接影響してくる可能性のある科学的・医学的知見を含むため、潜在的な社会的インパクトも大きく、コミュニティの「内外」を否応にも強く意識せざるを得ない

との意見が聴かれた [22]。

同じく冒頭で触れた英 *Lancet* 誌と米 *NEJM* 誌での論文撤回の事案からも示唆のあるとおり、「インパクトファクターの高い一流誌に掲載された論文だから質が保証されている」と受け取ることも、当然ながら必ずしも正しくない。査読付きジャーナル誌への掲載有無にもよることなく、研究者自身が常にアカデミック・インテグリティに基づく批判的精神と吟味の姿勢をもって個々の研究論文と向き合うことが益々欠かせない時代になっている。有識者からも、次々と新しいオープンアクセスジャーナルが発刊されて行く時代にあつて、研究情報の利用者にはジャーナルやそのエディトリアル質（玉石混淆）を適切に見極めていく努力が求められていくことが指摘された。その際、

プレプリントは元々物理学分野で始まったが、そこでは出版前のプレプリントであろうと、その内容が真っ当なものであるかどうか、考慮に値するものであるかどうかを読者側で自ら判断できることが前提のものとして活用されてきた。自分の頭で考え、既存の知識や言説を疑い、他に頼らず自ら評価を行う。そういうカルチャーの根付いた分野であればこそ、出版社の提供する査読サービスに過度に依存することなく、自律的かつスピーディーに研究を進めてこられた

との意見が聴かれた。いかにして質の担保された研究成果をオープンかつ迅速に公開・共有していくかという問題は、プレプリントかジャーナル論文かという形態の議論を越えて、いまや学術情報の公開に関してアカデミアが広く深く直面する課題となっている。科学的知見を生み出し発信していく研究者側、そして論文出版やプレプリントサーバーの運営等を通じてその発信活動を下支えする媒介者側にも努力が求められていく一方で、科学技術行政としても、政策検討にあたっては、そうした科学的知見がメディアを含む世の中に向けては必ずしも本来の意図そのままの形で伝わっていかない可能性があることを適切に織り込み、不断の努力を続けていく必要があるだろう。

3.3 日本の役割

各国においてプレプリントサーバーや各種リポジトリの整備が進められてきた中、日本でも各研究機関・FA や学協会等の単位を越えて国レベルでそうしたプラットフォームを整備することの意義については議論の聞かれるところである。今回話を伺った有識者の方々からは、アカデミアにとって真に必要なとされるものが行政主導ではなくアカデミアやコミュニティの主導で創られていくことが成功の必要条件との見解が共通して聴かれた上で、

仮に日本でプレプリントサーバーのようなものを持つにしても、世界中の研究者から見てそこがワンストップとして機能するくらいの一大大プラットフォームであって、過去の蓄積も最新情報も、全てのコンテンツがそこに集中しているくらいでないとうまくいかない。物理の世界で *arXiv* が成功したのは、そこに行けば全てがあるから。研究コミュニティが欲するものを自らの手で創り上げてきたから。いまから第二の「*arXiv*」を日本に持とうとするの意味は薄い。持つとするならば、日本固有の強み、日本文化に根ざしたようなものをコンテンツ化した上で、何より英語で発信していくことが必要

との意見が聴かれた [22, 23]。あわせて、

一時期、地域特化型のプレプリントサーバーが次々に立ち上がったがうまくいっていない。地域でコミュニティがよくまとまるということがあれば別だが、そうでない限り地域特化型を作ったところで機能しないだろう。例えばアジアでそのようなものができるか、ドメスティックなものに意味があるかという疑問。極端には、サーバーがどこにあろうと、分野で一個きちんとしたものがあればそれで十分。複数あったり地域限定であったりすることのメリットは少なく、2番手がやっても意味はない。ただ、例えばジャーナルにも採否基準があるのと同様に、プレプリントサーバーについても、欧米基準に対して何らか日本基準のものを用意するというのであれば話は別かもしれない。あるいは日本が特に強い分野で主導的に回せるプレプリントサーバーを持つというのであればあり得るかもしれない

との意見も聴かれた。加えて、

プレプリントサーバーを自前で持つことより、既成のものであろうとその中に最低限参加できているということが大事。ルールメイキングを主導したりコミュニティをリードしたりというところまでは行かないにしても。幸い、現状で *arXiv* のコミュニティには日本も入れている。*arXiv* 運営の会費の部分はアクセスの多い大学から集められているが、日本でも *NII* を中心に日本コンソーシアムとしてまとめて支払っている。欧米以外でこのコミュニティに入れている国は少ない中、このこと自体に価値がある

との意見も聴かれた [24]. さらに, こうしたプレプリントサーバーや研究情報のリポジトリを持続的かつ効果的に機能させていくに当たっては, arXiv が実際にそうであるように, 学協会との協調が欠かせないとの指摘も共通して聴かれた. 学協会が一丸となって外向きに積極的・戦略的に発信していくことが重要であり, 内向きに終始して学会誌自体が身内向けのサークル誌のようになっては広く求心力の伴った有用なプラットフォームとはなり得ないとの指摘もあった (cf. [42]). プレプリントサーバーの躍進を含め, 様々な状況がこの数年前の間でも大きく変化してきた中, 科学技術振興機構 (JST) の運営する電子ジャーナルプラットフォームである科学技術情報発信・流通総合システム (J-STAGE) [43] の将来的な立ち位置や役割についても意見が聴かれた. その中には,

リポジトリはプレプリントサーバーの代役とはならない. 日本の *J-STAGE* は, *arXiv* のように世界中の研究者が日々競って更新して戦っていく生き馬の目を抜くような戦場というより, これまでの学術知識の蓄積がユーザー目線で陳列されたある種の価値あるカタログ, 折に触れて歩き回ることによってセレンディピティを経験できるような図書館・博物館のような場所, 学術の多様性を守り活かしていく場所, そうした学術知識のアーカイブ機能を通じて社会に価値還元していくというのも一つの将来像ではなかろうか

との意見も聴かれた. 今後の我が国のデータ戦略に際して, arXiv 等のプレプリントサーバーとの相対的な関係で J-STAGE をどのように活かしていくかについても,

過去から未来まで全てカバーしないといけないと思うと大変だが, 過去はもう *arXiv* に任せれば良い. 未来に向けて今あるものをつなぐ. そのように新機軸を建てる. 論文やデータは DOI 等の ID, 研究者は ORCID 等の研究者 ID, 各種ソースもオープンならそういうものも含めてつないでいける. 日本語用の ORCID はないが, それも自動翻訳でなんとかできるのではないかな. 日本がやるとすれば日本自身や東アジア関係, そして日本が強みを有する材料系や, 非線形科学, 脳科学. そうした過去の上に立って様々なコンテンツをつないでいこうとするとき, これまでの *J-STAGE* の経験値も役立つのではないかな

との意見が聴かれた. 有識者からはまた別の観点として, arXiv データを活用することで, 日本ならではの視点で基礎研究支援の重要性を語る上でのエビデンスとしても活かしていくべきとの意見も聴かれた. ジャーナル上には表れない世界での活発な活動をどう捉えてどのようにエビデンスとして活かしていくかについて, 有識者からは,

教科書を書き換えるような新しいアイデアの源泉を見出し評価していけるよう, 長期スパンでの評価指標で見ていくべき. 日本が過去に果たしてきたまだ知られぬ学術的貢献で, いまや世の中に大きな価値を生んでいるものが過去のデータの中に眠っている, そのことのエビデンスが '見える化' できると良い. 科研費を通じて, 欧米に比べて長いスパンでの研究支援を続けてきたことが日本の強みのはず. 数十年後になって日の目を見たような研究成果は探せばたくさん出てくるはずで, それをなんとか見出して育ててほしい. 基礎研究がいかに社会を支える土台を作っているか, 行政側で事例を掘り起こしてほしい

との意見が聴かれた [22]. また, これまでのジャーナル論文情報 (大手学術出版社の提供するデータベースに掲載のあるもの) のみをもとに科学技術政策上の投資先を論じていくことは, 既に数年前に終わったことの実績を確認する行為であり, '終わったもの' にリソースを割いていく政策になりがちだが, プレプリントサーバー上の情報を分析することで, まだ萌芽的な段階で「いま来ている・立ち上がりつつある」という新たな研究進展の兆しを捕捉していける可能性があるとの意見が聴かれた. 具体的には,

arXiv のデータからプレプリントの連結度等の分析を通じて、「ここに人が集まりつつある・インパクトが生まれつつある」という兆しを捉えるような実績を見せられると良い。いまの日本の研究の強みを知ることにもなる。既に人の集まった場所ではなく、「(集まっている場所の) 間をつないでいく」という発想とアプローチが重要。そこにファンドを付けていくような流れができていくと良くなるのではないか。プレプリントのデータも含め全てをつないでいくことでこれまでにない価値が生まれていく

との意見が聴かれた。オープンサイエンス時代にあって、出版社ビジネスとしては従来の論文出版サービスだけではなく種々のインフラ整備こそが今後の成長基盤となるだろう。欧米の大手出版社側でも新たな研究サービスも次々と生み出されている中、有識者からは、

欧州は新たに各種 ID (研究者や論文) を紐付ける研究サービスを作ろうとしており、このような‘種を蒔く’イノベーション・サービスに長けている。このようなサービスは日本単独で手を出すのは難しいかもしれないが、戦略的に欧州と組むというのも手だろう。でき上がったものに乗るのではなく、できる前から関与することが大事。様々な研究インフラのルール作りの国際会議を見ていると日本のプレゼンスは非常に低い。今後そのような研究インフラに投資していくことは価値があるのではないか

との意見が聴かれた。冒頭でエルゼビア社による SSRN 買収の件に触れたが、ジャーナル発行によりその購読料や APC (論文掲載料) 等で利益を得ていた大手学術出版社も、オープンサイエンス時代を生き延びていく上で、いまやプレプリントの存在を織り込んだ、あるいは積極的に利用した新たなビジネス形態に舵を切らざるを得ない時代になっている。確かにプレプリントサーバー自体は直接的な収益源とはならないが、プレプリントの段階から関与して研究者や研究論文を青田刈りしておけば自らのジャーナルビジネスにもつながるといふ算段と言われる [24]。より大局的な見地から官民の役割の在り方を考える上では、有識者からは *arXiv* を例にとりつつ、

プレプリントサーバーでは運営側の持ち出しが多くなる。*arXiv* も 3 分の 2 は米国の資金。会費分も半分は米国で、したがって都合 6 分の 5 は米国の持ち出しとなっている。これは自分たちが強い立場なのだから先導して支援する、そのための支援は惜しまない、ということ。金を出しても口は出さないというのはアカデミアへの正しい関与姿勢。そして必ずしも政府が主導するわけでもない。大学やアカデミアが自分たちで決めて運営するという米国の自主性文化の顛れが *arXiv* の運営にも見て取れる

との意見が聴かれたが、このことも今後の我が国としてのオープンサイエンス時代を官民協働で、そして国際協調・競争の下で進めていくに当たって示唆に富む。

3.4 おわりに

学術情報のコミュニケーションの在り方は時代とともに大きく変遷してきた。かつては個人的な研究交流 (郵便、電話、FAX) やいわゆる灰色文献、非公式なセミナーを含む研究集会やカンファレンス、そしてよりフォーマルなジャーナルや書籍といった形であったものが、情報通信技術の飛躍的な進展に伴い、電子メール、電子メーリングリスト、プレプリントといった新しいコミュニケーションの形が従来のものを補完ないし代替してきた。さらに近年では各種の SNS やブログ等のソーシャル・メディアが幅広く活用され、今般のコロナ禍ではリモート会議のための様々な環境整備やコンテンツの充実化もなされ、いまや研究コミュニティにとって欠かせない研究インフラとなっている。学術情報の在処や流通手段は、いまや個人の物理空間を大きく越えてクラウド上のデジタル空間へと不可逆に移行している。それはほんの一年前には想像もつかなかったほどのスピードで日々進化している。今後、プレプリントやそれ以降の新しい学術情報の流通手段が変えていくアカデミアや世の中の姿とは一体どのようなものであろうか。膨大な量の学術情報が広くアカデミア内外を、必ずしもそれ自体には質保証の伴わない形で駆け巡っていく時代にあって、科学と社会との在るべき関係性を描いていく上で我々は何を道標としていくべきであろうか。そして科学技術行政は

そうした新たな時代をどのように先導していけるであろうか。

目下 NISTEP において、日本国内の研究機関に在籍する研究者等を対象としたプレプリントの活用状況や意識等に対するアンケート調査 [32] が行われているが、その結果からも、プレプリントの利用実績や定着度は分野によって大きく異なることが示唆されている。冒頭で述べたプレプリントの Pros や Cons の観点やその程度も分野によって大きく異なる。その背景にあるものは様々な歴史的経緯であり、分野カルチャー¹⁷であり、また影響力の大きな学術出版社のスタンスに至るまで、広範かつ多岐にわたるだろう。今後の科学技術政策への反映に当たっては、こうした数々の分野特性にも十分に留意して上でさらに検討が深められていく必要がある。各分野でダイナミックな進展を遂げてきた、そしてさらに発展を続けているプレプリントサーバーと日々向き合っていくことが、行政の立場からも一層欠かせない時代となっている。プレプリントに留まることなく、新しい技術の進展やアイデアの導入に伴って研究者による知の生産活動や学術情報の流通の在り方は今後も様々な要因（例えば分野の差、世代の差、国や地域の差）から益々チャレンジングな時代を迎えていくだろう。

本稿において現代的な視点から改めて整理・再考したプレプリントの意義や在るべき位置付け、そして実際の arXiv データをもとに得られた科学発展への寄与についての定量的なエビデンスが今後の科学技術政策に活かされていくことを期待したい。また、研究者や政策関係者のみならず、広くアカデミア、行政、政治、民間、そして科学ジャーナリズムを含む関係者間でさらなる調査・分析や議論が行われる際の検討材料として活かされていくなら幸いである。

謝辞

本稿の執筆にあたり御知見・御助言をいただいた野崎光昭氏（高エネルギー加速器研究機構; KEK）、引原隆士氏（京都大学）、武田英明氏（国立情報学研究所; NII）に御礼申し上げる。各氏からは、本稿のテーマであったプレプリントに関する話題に留まらず、より広く学術情報発信の在り方、研究評価の在り方、国際戦略、オープンサイエンスのさらにその先の展望を含め、様々な観点から御意見をいただいた。本稿で紹介しきれなかった話題や掘り下げきれなかった内容については、arXiv 等のデータをもとにした追加的な分析結果等とも合わせて、また稿を改めて論じたい。また、数々の有用な御意見・フィードバックをいただいた上山隆大氏（総合科学技術・イノベーション会議; CSTI）、鈴木了氏（Shing-Tung Yau Center of Southeast University）、井出和希氏（京都大学）の各氏に加え、文部科学省及び科学技術・学術政策研究所（NISTEP）の関係者の皆様に重ねて御礼申し上げる。なお、本稿に示された見解や記された内容は著者ら個人によるものであり、著者らの所属する組織を代表するものではない。

参考文献

- [1] *Publish or perish*. https://en.wikipedia.org/wiki/Publish_or_perish.
- [2] Kendall Powell. Does it take too long to publish research? *Nature*, 530:148–151, 2016.
- [3] Daniel Johnston. *Some Things Last A Long Time*. <http://quantixed.org/2014/04/15/some-things-last-a-long-time/>.
- [4] Heather A. Piwowar, Jason Priem, V. Larivière, J. P. Alperin, L. Matthias, Bree Norlander, Ashley Farley, J. West, and Stefanie Haustein. The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, 6, 2018.
- [5] Jonathan P. Tennant, Harry Crane, Tom Crick, Jacinto Davila, Asura Enkhbayar, Johanna Havemann, Bianca Kramer, Ryan Martin, Paola Masuzzo, Andy Nobes, Curt Rice, Bárbara Rivera-López, Tony Ross-Hellauer, Susanne Sattler, Paul D. Thacker, and Marc Vanholsbeeck. Ten hot topics around scholarly publishing. *Publications*, 7(2):34, 2019.
- [6] arXiv.org [Preprint server]. <https://arxiv.org/>.
- [7] Paul Ginsparg. Preprint déjà vu. *The EMBO Journal*, 35(24):2620–2625, 2016.

¹⁷研究分野によって異なる学会やジャーナルの作法、分野特性を反映した論文媒体・流通手段（‘TeX vs. Word’）、研究スタイル（‘Dry vs. Wet’）、研究不正の起こりやすさ等がプレプリントの活用のされ方や浸透度、研究コミュニティ内での受け止めに大きく影響するものと考えられる。

- [8] Philip E. Bourne, Jessica K. Polka, Ronald D. Vale, and Robert Kiley. Ten simple rules to consider regarding preprint submission. *PLoS Computational Biology*, 13, 2017.
- [9] Sarvenaz Sarabipour, Humberto J. Debat, Edward Emmott, Steven J. Burgess, Benjamin Schwessinger, and Zach Hensel. On the value of preprints: An early career researcher perspective. *PLOS Biology*, 17(2):1–12, 2019.
- [10] Transpose database. <http://transpose-publishing.github.io/>.
- [11] *List of academic journals by preprint policy* [Webpage]. https://en.wikipedia.org/wiki/List_of_academic_journals_by_preprint_policy.
- [12] What are preprints? *YouTube video produced by ASAPbio and Youreka Science*, 2016.
- [13] Michael S. Lauer, Harlan M. Krumholz, and Eric J. Topol. Time for a prepublication culture in clinical research? *The Lancet*, 386(10012):2447–2449, 2015.
- [14] Ronald D. Vale. Accelerating scientific publication in biology. *Proceedings of the National Academy of Sciences*, 2015.
- [15] bioRxiv [Preprint server]. <https://www.biorxiv.org/>.
- [16] ChemRxiv [Preprint server]. <https://chemrxiv.org/>.
- [17] medRxiv [Preprint server]. <https://www.medrxiv.org/>.
- [18] Naomi C. Penfold and Jessica Polka. Preprints in biology as a fraction of the biomedical literature (Version 1.0) [Data set], 2019.
- [19] Naomi C. Penfold, Fiona L.M. Murphy, and Jamie J. Kirkham. Practices and policies of preprint platforms for life and biomedical sciences (Version 2.0) [Data set], 2020.
- [20] 伊藤 裕子, 小柴 等, 林 和弘. COVID-19/SARS-CoV-2 関連のプレプリントを用いた研究動向の試行的分析. *NISTEP Discussion Paper*, 186, Jun 2020.
- [21] 池内 有為. オープンサイエンスの効果と課題—新型コロナウイルスおよび COVID-19 に関する学術界の動向. *情報の科学と技術*, 70(3):140–143, Mar 2020.
- [22] 野崎 光昭. Personal communication, 2 Sep. 2020.
- [23] 引原 隆士. Personal communication, 3 Sep 2020.
- [24] 武田 英明. Personal communication, 7 Sep 2020.
- [25] Prashant Pradhan, Ashutosh Kumar Pandey, Akhilesh Mishra, Parul Gupta, Praveen Kumar Tripathi, Manoj Balakrishnan Menon, James Gomes, Perumal Vivekanandan, and Bishwajit Kundu. Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag. *bioRxiv*, 2020.
- [26] Tom Sheldon. Preprints could promote confusion and distortion. *Nature*, 559:445, 2018.
- [27] Jonathan Tennant. Preprints help journalism, not hinder it. *Nature*, 560:553, 2018.
- [28] Clarissa F. D. Carneiro, Victor G. S. Queiroz, Thiago C. Moulin, Carlos A. M. Carvalho, Clarissa B. Haas, Danielle Rayêe, David E. Henshall, Evandro A. De-Souza, Felipe E. Amorim, Flávia Z. Boos, Gerson D. Guercio, Igor R. Costa, Karina L. Hajdu, Lieve van Egmond, Martin Modrák, Pedro B. Tan, Richard J. Abdill, Steven J. Burgess, Sylvia F. S. Guerra, Vanessa T. Bortoluzzi, and Olavo B. Amaral. Comparing quality of reporting between preprints and peer-reviewed articles in the biomedical literature. *bioRxiv*, 2020.
- [29] Rob Johnson, Anthony Watkinson, and Michael Mabe. The STM report: An overview of scientific and scholarly publishing (5th ed.). Technical report, STM: International Association of Scientific, Technical and Medical Publishers, October 2018.
- [30] Darwin Y. Fu and Jacob J. Hughey. Meta-Research: Releasing a preprint is associated with more attention and citations for the peer-reviewed article. *eLife*, 8:e52646, 2019.
- [31] Jaime A. Teixeira da Silva. The preprint wars. *AME Medical Journal*, 2(6), 2017.
- [32] 林 和弘, 池内 有為. (プレプリントに関する NISTEP の調査) . *NISTEP Discussion Paper*, 2020.
- [33] Smriti Mallapaty. Popular preprint servers face closure because of money troubles. *Nature*, 578:349, 2020.

- [34] SSRN (Social Science Research Network) [Preprint repository]. <https://www.ssrn.com/index.cfm/en/>.
- [35] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: Bringing order to the Web. *Proceedings of the 7th International World Wide Web Conference*, pages 161–172, 1998.
- [36] 孫 媛. 引用に基づく学術研究のインパクト評価. *情報の科学と技術*, 70(5):255–260, 2020.
- [37] 林 和弘, 小柴 等. arXiv に着目したプレプリントの分析. *NISTEP Discussion Paper*, 187, 2020.
- [38] *arXiv submission rate statistics*. Data for 1991 through 2019, updated 1 January 2020. https://arxiv.org/help/stats/2019_by_area/index.
- [39] 鷺尾 隆. 一流国際会議発表のための研究戦略とは? . *人工知能学会誌*, Vol.23, No.3, pp.362–366, May 2008. <http://id.nii.ac.jp/1004/00006982/>
- [40] 住井 英二郎. 「情報系」の業績評価について—「若手」研究者の視点から—, 2019. 日本学術会議 科学者委員会 研究評価分科会 公開シンポジウム 「研究評価の客観化と多様化をめざして—分野別研究評価の現状と課題」.
- [41] ASAPbio — *Funder policies*. <https://asapbio.org/funder-policies/>.
- [42] 日本学術会議第三部理工系学協会の活動と学術情報に関する分科会. 学術情報流通の大変革時代に向けた学術情報環境の再構築と国際競争力強化, 2020.
- [43] J-STAGE [electronic journal platform]. <https://www.jstage.jst.go.jp/browse/-char/ja>.