

レポート

COVID-19 で加速するオープンサイエンス
— プレプリント分析にみる学術情報流通の変容 —

科学技術予測センター 上席研究官 林 和弘

概要

2021年時点のオープンサイエンスの進展について、プレプリントを中心とした動向として、科学技術・学術政策研究所（NISTEP）で行った分析、意識調査を用いて解説する。プレプリントは、COVID-19によってその価値が幅広く認識されるだけでなく、研究動向分析の新たな手段を提供している。また、プレプリントの共有は既に一部の分野では30年近くをかけて浸透しており、arXivの分析によって、情報学ではプレプリント引用の慣習が生まれていることや、arXiv関連論文の総被引用数の4割以上をプレプリントが関連する被引用で占められていることが示された。一方、プレプリントの受容とその将来展望は世代間や分野による差があることも意識調査から分かった。プレプリントの取扱いは依然慎重に議論されるべきものではあるが、その特性を生かして、研究者に新たな成果公開の戦略的手段を与えるものであり、原著論文とその分析を補完するものとしても活用されるべきものである。欧米でもプレプリントの扱いが変わる中、第6期科学技術・イノベーション基本計画では、研究のデジタルトランスフォーメーションやオープンサイエンスがプレプリントとともに取り上げられる方向にある。

キーワード：オープンサイエンス、プレプリント、COVID-19、arXiv、学術情報流通

1. はじめに

オープンサイエンスは、政策を中心に公的資金を得た研究論文と研究データを主とした研究成果のよりオープンな共有・公開によるイノベーションの加速と社会の変容を指向してきた¹⁾。日本でもオープンサイエンス政策が本格的に始まった2014年頃より2019年までは、どちらかと言えば、将来に起こりうる予察に対する準備や先行者利益獲得の意味合いが強かった²⁾。しかしながら、COVID-19という世界全体の危機的な社会課題に対応するために学術知が迅速に共有されることによって^{3,4)}、図らずもオープンサイエンスの意義が具体的かつ実効を伴って喫緊の課題として再確認されることとなった。特にプレプリントの共有による迅速な成果の公開と共有によって、学術情報のオープン化の必要性が広く認識されるだけでなく、既存の査読や学術ジャーナルの在り方、あるいは知財の在り方も問い直されている。そして、研究者とそれらを取り巻くあらゆるステークホルダー

が、新たな学術知の発見と共有に向けた実践に取り組んでおり、オープンサイエンスのビジョンが持つ科学研究そのもののデジタルトランスフォーメーションに向かいつつある。

その一方で、社会構造の変革も伴う変化の過渡期においては、オープンサイエンスの推進には分野間あるいは世代間の意識の差等に象徴される課題も多数存在し、それらを変化の駆動要因と合わせてできるだけ正しく把握し、その結果を踏まえた変革を促す方策をとることが政策においても重要である。

本レポートでは、プレプリントが進展する兆しを報告した前報⁵⁾を踏まえて2021年時点におけるオープンサイエンスの現状を、科学技術・学術政策研究所（NISTEP）で行ったプレプリントに関する分析や意識調査を通じて解説し、オープンサイエンスを更に進めるための課題について議論する。

2. プレプリントの共有・公開とその分析から明らかになったこと

COVID-19 は前述の通りプレプリントの迅速な共有を促したが、その現象及び、その影響をある程度定量的に把握する必要があり、そのことでEBPM (Evidence Based Policy Making: エビデンスに基づく政策立案) に寄与することが可能となる。

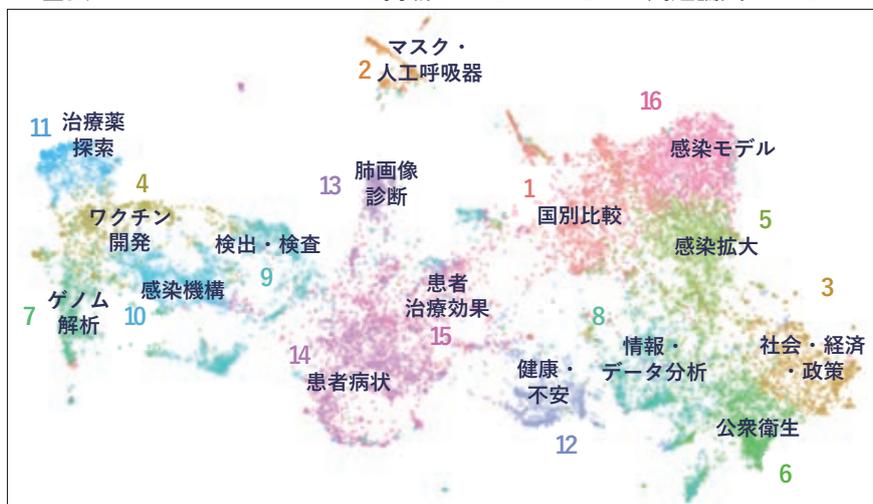
NISTEP では、まず、arXiv や medRxiv など主要なプレプリントサーバにおける COVID-19 に関するプレプリントの概況把握を定量的に試みた⁶⁾。その結果、全体として COVID-19 に関するプレプリント投稿件数が伸びていることや、医療系のみならず、人文社会系や物理・情報系に主軸を置くプレプリントサーバにおいてもプレプリントの投稿があることなどが確認され、幅広い分野の研究者の貢献が認められた。加えてその論文の内容 (トピック) について、自然言語処理を用いて分類したところ、通常の査読論文も含めた同様の分析では明確には検出ができていなかった、医薬・ワクチン開発に関するトピックを抽出することができた。このように、プレプリントが浸透し、その群を対象にデータ処理をすることによって、エマージングな研究動向を把握できる可能性など、原著論文と被引用数を用いた分析より早く研究の見える化 (Visualization) を行うことが可能となった (図表 1)。

COVID-19 に関する調査は、いわば、緊急対応下におけるプレプリントの現状把握とも言えるが、より一般的な傾向を分析することを目的に、プレプリントを 30 年近く運用している arXiv に掲載された 160 万本を超えるプレプリントを分析した⁷⁾。その結果、物理学、情報学を中心にプレプリントが広く受け入れ

られているだけでなく、情報学においてはプレプリント投稿の大幅な伸びを確認し、プレプリントを研究成果物として活用し、引用する慣習も示唆される結果となった。また、プレプリントの 1/3 程度は、原著論文や書籍等の既存の成果公開と認められる出版物になっていないことも明らかになった。これらは、定性的にはその分野の研究者で主張されてきたことであり、特に情報学の分野では、日進月歩のために、ときに数か月以上かかる査読や原著論文の出版を待つことなくプレプリントの共有が行われているとされているが、それをある程度定量的に裏付けることができた。すなわち、COVID-19 関連のプレプリントと論文が共有される過程で、査読による質の担保が注目され、その重要性が再認識されつつも査読の課題が改めて浮き彫りになったが、このような査読が持つ構造的な問題は COVID-19 より前に顕在化しており、分野依存ながらプレプリントによる代替手段も発展していたと言える (図表 2)。

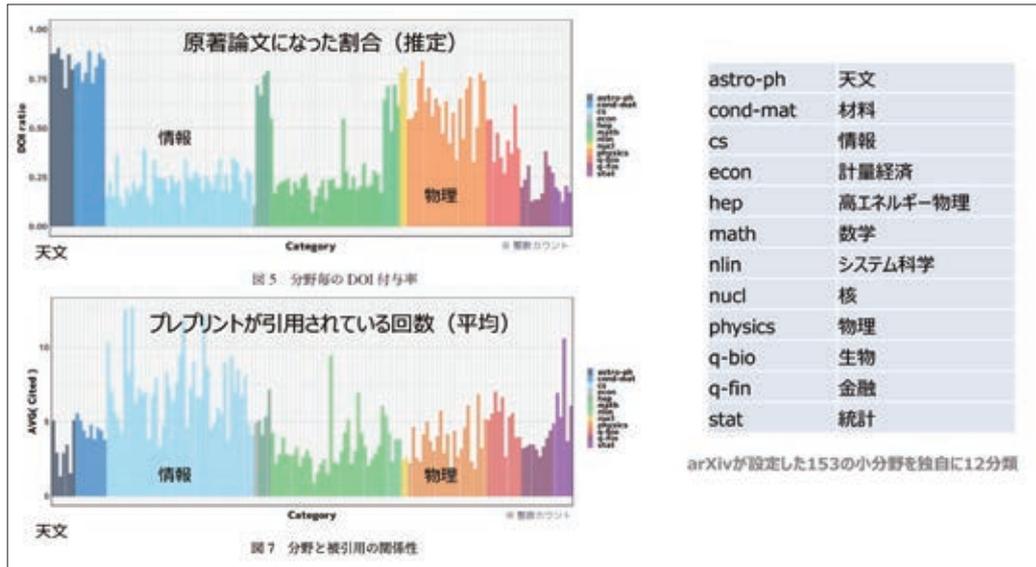
さらに、NISTEP が文部科学省関連部局 (職員) と共同で arXiv に関して更に詳細な分析を行った結果⁸⁾、プレプリントが既に浸透しているとされる、数学や情報学の存在を再確認した。さらに、総被引用回数におけるプレプリント関連の被引用数の割合 (論文と被引用によるネットワークにおけるプレプリントの存在感) を算出したところ、新規投稿からしばらく年月が経つと、先行している数学や情報学において、その割合が 6-7 割に達していることに加えて、その他の分野においても、約 4 割がプレプリント関連の被引用数で占められていることが分かった (図表 3)。このプレプリントの引用ネットワークにおける存在感と先の arXiv の 1/3 程度は既存の出版物にならない結果からは、研究成果の多様性とその多様な成果を

図表 1 プレプリントサーバに掲載された COVID-19 関連論文のトピック



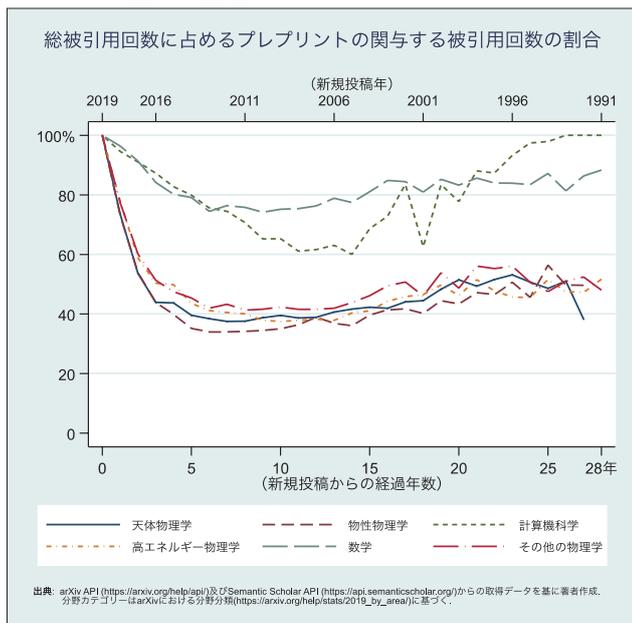
出典：参考文献 6)

図表 2 arXiv の掲載プレプリントの分野別分析



出典：参考文献7)

図表 3 arXiv 上のプレプリントによる被引用が総被引用に占める割合



出典：参考文献8)

活用した研究活動の存在が示唆され、原著論文の分析だけでは見えていなかった研究活動を付加的に把握できる可能性をある程度定量的な議論の上で示すことができた。

以上のように、プレプリントは、COVID-19 によって短期的に幅広く注目されるようになって研究動向把握の新たな手段を我々に与えただけでなく、先行している分野においては、研究者の研究活動と研究成果の積み重ねによる科学の発展において、既に無視できない存在になりつつあることが明らかとなっている。

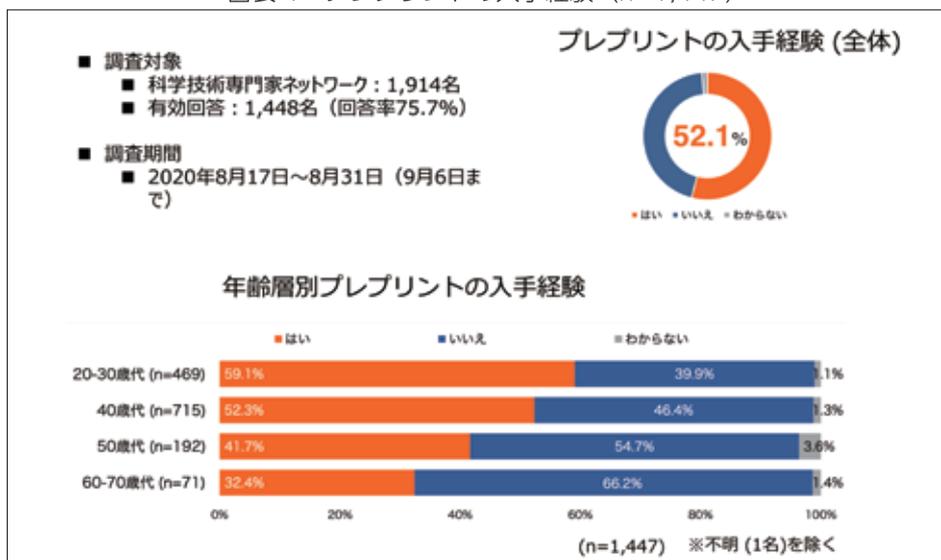
また、プレプリントは、飽くまで論文の草稿であり、質の担保に課題を抱えながらも、オープンサイエンスが指向する科学のデジタルトランスフォーメーションについて現在の関係者に理解しやすいイメージを提供していると言える⁹⁾。このことは、オープンサイエンス政策の柱である研究データの共有・公開において、その具体化の難しさやインセンティブとの関連が乏しいという現状と比較すると、研究者及びそのコミュニティと関係者に変容を促す上で大きな利点である。

3. 日本の科学技術専門家の意識調査にみる予察

プレプリントの一定のインパクトや可能性を計量書誌学的な分析で認めた上で、NISTEP では、今後の学術情報流通政策に資するために、2020年8月から9月にかけて日本の研究者によるプレプリントの利活用の状況と認識に関するオンライン調査を実施した¹⁰⁾。対象は科学技術予測センターが運営する科学技術専門家ネットワークであり、1,448名から回答を得た(回答率75.7%)ものである。

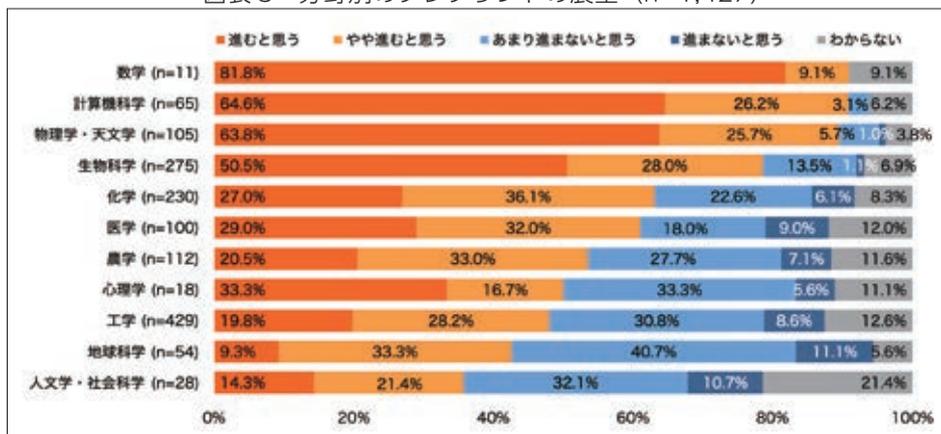
その結果、プレプリントの受容は年代別に差があることが分かり、今後の変容を示唆するものとなっている(図表4)。また、プレプリントの受容やその展望に、分野別に差があることも分かった。展望の分野別の差については、図表に示す通り、計算機科学、物理学・天文学、生物科学等プレプリントサーバが進展している分野が今後も進むと考えている割合が高く、工学、地球科学、人文学・社会科学では、その割合が低

図表 4 プレプリントの入手経験 (n=1,447)



出典: 参考文献 9)

図表 5 分野別のプレプリントの展望 (n=1,427)



出典: 参考文献 10)

いことが分かった (図表 5)。

4. プレプリントの位置づけと政策

以上の定量的な分析結果や意識調査から、プレプリントの位置づけを政策の面から考察する。まず、プレプリントを研究成果として考える際に、プレプリントは原著論文の代替物とはならないことと、付加的な研究成果として、原著論文ではできない分析を補完的に行える関係にあることを認識することが重要である。これは、逆にプレプリントが役に立たないということでもなく、より早く研究成果の候補を把握する、あるいは、原著論文の群の調査とはまた違った価値観から研究動向を把握するためには有用と考えるべきである。

続いて、分野によってプレプリントの扱いや意識が異なる点には、改めて注意が必要であり、施策を講ず

る上でも今しばらくは分野別の対応が求められる。むしろ、現時点では、幅広い分野において研究者の研究成果公開戦略において新たな手段が加わったと考えられる。すなわち、分野を問わず、個々の研究ごとに先取権の確保に加えて自身の研究成果を広く認めてもらうことや研究者としてのキャリア形成において有利になる場合、あるいは今回の COVID-19 のように早急な対応が必要で、また分野横断協業が求められる場合はプレプリントを有効活用することが考えられる。そして、プレプリントの有効活用が幅広い分野で慣習化することを推進する施策は公的資金を用いた研究成果の迅速な共有の観点から検討に値する。

一方、読者側としてみた場合は、読み手の専門性や立場に応じてプレプリントを使い分ける必要がある。例えば、自身の専門性から、プレプリントで早く情報を入手して、自身の研究等に役立てること自体は何の問題もない。その一方で、プレプリントのすべてを査

読済み原著論文と同等の質が担保されているものとして扱うことはできず、研究機関等組織として研究成果を取り扱う場合、特に広報する場合には注意が必要である。

また、個々の研究成果としてプレプリント自体を認めるかどうかは分野によって大きく分かれており、先の調査でも、プレプリントが評価や昇進につながることを示唆する結果も部分的に得られている。研究助成団体を中心に、プレプリントの成果としての取扱いについては依然慎重な議論と運用が求められるが、例えば査読中の論文のプレプリントを必要に応じて何らかの成果として登録できる仕組み自体は予め整えておき、分野やタイミングに応じて臨機応変に運用できるようにしておくことは考慮に値する。

プレプリントの成果としてどう位置づけるかについては、欧米でも試行錯誤が行われている。例えば、欧州でも PMC Europe において、2018 年以降 22 万本以上プレプリントを公開している¹¹⁾。米国では、NSF において 2018 年より原著論文として出版されることが確定したプレプリントを一部受け付けることを表明しているが¹²⁾、最近では、米国国立衛生研究所 (NIH) において、プレプリントパイロットプログラムが立ち上がり、NIH が研究助成した成果のプレプリント登録が始まった¹³⁾。

日本においては、第 5 期科学技術基本計画よりオープンサイエンスの推進を掲げて、内閣府や G7 科学技術大臣会合、及び文部科学省において検討が進められてきた。そして、第 6 期科学技術・イノベーション計画に向けた検討の案¹⁴⁾ において、“コロナ関連の研究のグローバルな発信を契機に、研究成果の共有の仕組みとして、プレプリントサーバの活用の動きが

存在感を増してきており、各国でオープンアクセス、オープンサイエンスのプラットフォームづくりの動きが盛んになっている。”とし、研究のデジタルトランスフォーメーションとともに、オープンサイエンス、プレプリントも取り上げられており、具体的な施策に落とし込むことが見込まれている⁹⁾。

4. おわりに

オープンサイエンスが目指す長期的ビジョンは、インターネットを用いた情報基盤の根本的な変革に基づく科学と社会及び“科学と社会”の根本的な構造変革¹⁾ であり、その変革を駆動する一つの要素として、プレプリント、原著論文、研究データのよりオープンな共有と公開が位置づけられる。そして、現在は研究成果公開の新しいメディアが様々に試行されており、その信頼性のある程度時間をかけて確保した後に、そのメディアの活用が慣習として受け入れられることになる。学術情報流通の観点からは、この過程を経た上で、原著論文とその引用関係を補完・拡張する新たな研究活動の生態系とその系に基づく評価手法を生み出すことが予察される。このような研究者コミュニティ自身の変革を伴う変化、中でも評価の変容については、研究者による自発的な変化だけで行うのは難しいと考えられる。研究者の主体性自身は決して損なわれるべきものではないが、外部からの変化の刺激を与えることが重要であり、引き続き政策が果たす役割は重要である。そして、その政策づくりと実装においてもデジタルトランスフォーメーションが必要¹⁵⁾ であると考えられる。

参考文献・資料

- 1) 林 和弘. オープンサイエンスの進展とシチズンサイエンスから共創型研究への発展. 学術の動向. 2018, vol.23, no.11, pp.12-29. https://doi.org/10.5363/tits.23.11_12
- 2) 統合イノベーション戦略におけるオープンサイエンス—研究データの戦略的開放による「知の源泉」を担う基盤づくりに向けて—. STI Horizon. 2018, vol.4, no.3, p.42-47. <https://doi.org/10.15108/stih.00145>
- 3) 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. 2020, vol.70, no.3, p.140-143. https://doi.org/10.18919/jkg.70.3_140
- 4) 尾城孝一. 進展するプレプリントの風景. 情報の科学と技術. 2020, vol.70, no.2, p.83-86. https://doi.org/10.18919/jkg.70.2_83
- 5) 林和弘. MedRxiv, ChemRxiv にみるプレプリントファーストへの変化の兆しとオープンサイエンス時代の研究論文. STI Horizon. 2020, vol.6, no.1, p.26-31. <https://doi.org/10.15108/stih.00205>
- 6) 小柴等, 林和弘, 伊藤裕子. 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. 文部科学省科学技術・学術政策研究所, 2020, NISTEP DISCUSSION PAPER No.186, 10p. <https://doi.org/10.15108/dp186>

- 7) 林和弘, 小柴等. arXivに着眼したプレプリントの分析. 文部科学省科学技術・学術政策研究所, 2020, NISTEP DISCUSSION PAPER No.187, 24p. <https://doi.org/10.15108/dp187>
- 8) MEXT-NISTEP プレプリント調査・検討チーム. 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. https://www.mext.go.jp/content/20201026-mxt_jyohoka01-000010684_2.pdf
- 9) 林和弘. COVID-19 で加速するオープンサイエンスと政策. 第35回研究・イノベーション学会年次学術大会講演要旨. 35(1C06).
- 10) 池内有為, 林和弘 「プレプリントの利活用と認識に関する調査」, NISTEP RESEARCH MATERIAL, No.301, 文部科学省科学技術・学術政策研究所. <https://doi.org/10.15108/rm301>
- 11) Preprints - About - Europe PMC europepmc.org/Preprints
- 12) NSF. Proposal & Award Policies & Procedure Guideline. Chapter V - Renewal Proposals https://www.nsf.gov/pubs/policydocs/pappg18_1/pappg_5.jsp
- 13) NIH. NIH Preprint Pilot. <https://www.ncbi.nlm.nih.gov/pmc/about/nihpreprints/>
- 14) 科学技術・イノベーション会議 基本計画専門調査会 科学技術・イノベーション基本計画の検討の方向性 (案) <https://www8.cao.go.jp/cstp/tyousakai/kihon6/chukan/>
- 15) 林和弘, 吉本陽子, 佐藤遼, 鈴木羽留香. デジタルイノベーションとイノベーション政策. 研究技術計画. Vo.34, No.3, pp.270-283 (2019) https://doi.org/10.20801/jsrpim.34.3_270