

第 2 回 SPARC Japan セミナー2020

「プレプリントは学術情報流通の多様性をどこまで実現できるのか？」

コロナ時代における研究情報発信 - プレプリントに関する出版社の取り組み

アントワン・ブーケ

(シュプリング・ネイチャー (日本))

講演要旨



研究成果への早期かつ迅速なアクセスは、科学のペースを加速させ、発見を進展させるために重要である。特にコロナ時代において、その重要性は顕著である。そのため、出版社は、研究をできるだけ早くアクセスできるように取り組んでいる。シュプリング・ネイチャーのジャーナルは、長きにわたって、プレプリントの共有をサポートしてきた。また、Nature のジャーナルおよび BMC ジャーナルは、20 年以上も前から、出版前の原稿を共有することへの価値を認識してきた。

シュプリング・ネイチャーでは、ジャーナルへの投稿と査読（ピアレビュー）プロセスに、早期共有と透明性の高い査読を統合した革新的なサービス、In Review（イン・レビュー）を導入している。

今回の発表では、コロナ時代における研究出版の変化、プレプリントの長所と短所、および In Review を通じてどのように研究の早期共有を推進しているかを紹介する。また、どのようにして研究をオープンにしているのか、さまざまなオプションについても解説する。

アントワン・ブーケ

アントワン・ブーケは、シュプリング・ネイチャーのヴァイス・プレジデントとして、東京を拠点に日本、東南アジア、オセアニアのインスティテューショナル・セールスを担当している。アジア太平洋地域の学術出版界で20年以上の経験を持ち、シュプリング・ネイチャー（日本）の代表取締役社長も務める。これまでアジア地域の出版プログラムを指揮したほか、日本で医学情報事業を立ち上げた。また、書籍のコミッショニング・エディターとしての経験もある。オーストラリアで生まれ、ブリスベンのグリフィス大学卒業。東京大学で博士号（物理学）を取得。1994年から日本在住。



COVID-19 に関連する研究出版物

本日紹介するデータは、ほとんどが Digital Science の Dimensions のデータベースから 2020 年 10 月ごろに取り込んだものですが、Digital Science の COVID-19 のデータベースはリアルタイムで毎日アップデートされているので、そのデータレポートも紹介します。また、2020 年 6 月に Digital Science が COVID-19 の研究成果に関する大変面白いレポートを発表したので、その詳細もぜひご覧ください。

Dimensions のデータを分析すると、2020 年 1 月か

ら、COVID-19 に関連する 21 万 5,000 報以上の研究出版物が 1 万 7,500 機関から発表されています（図 1）。そのうち 3 万 4,000 件（17%）はプレプリントで、6,000 機関近くから発表されています。特に各国のロックダウンなどが始まると、4 月、5 月ごろから研究論文とプレプリントの数が大幅に増加しました。研究論文は 10 月まで毎月ほとんど同じくらいの程度で出版が続き、プレプリントはだんだん減っています。COVID-19 に関する研究論文のトータル数は、Dimensions データベースに入っている研究論文の約

3.6%で、医療・保健分野の約 14%を占めます。毎年、世界の研究論文は 2~3%成長するといわれているので、COVID-19 に関する研究はその成長を 2 倍ぐらい加速させたと言えるかもしれませんが、このコロナ時代の間に、COVID-19 に関する研究だけでなく、全体的に研究のアウトプットが大幅に増えました。

COVID-19 に関連する研究出版物のうち、17%はプレプリントです。同じ検索カラムで 2019 年のデータを見るとたった 6,000 件だったので、ほとんどは今年に出版されたものです。

COVID-19 に関するプレプリント

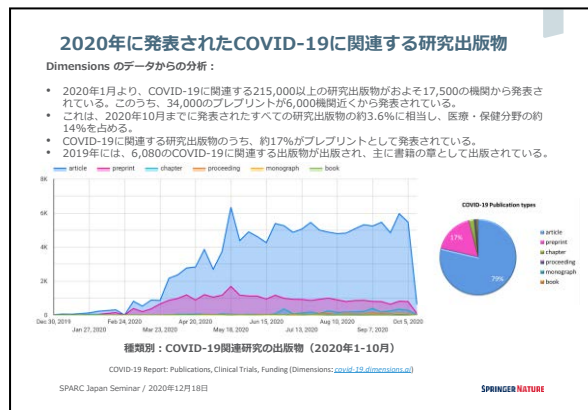
世界の COVID-19 に関連するプレプリントは、世界の COVID-19 に関連する出版物のうちの 17%です（図 2）。同じ時期に出版されている研究出版物全体に占めるプレプリントの割合はたったの 6%なので、COVID-19 に関する研究のプレプリントはその 3 倍で

す。特に中国の研究者は約 4 分の 1 がプレプリントの記事で出版しています。日本の場合は 19%です。そして、医療・保健分野のプレプリントのうち 40%が COVID-19 に関するものでした。

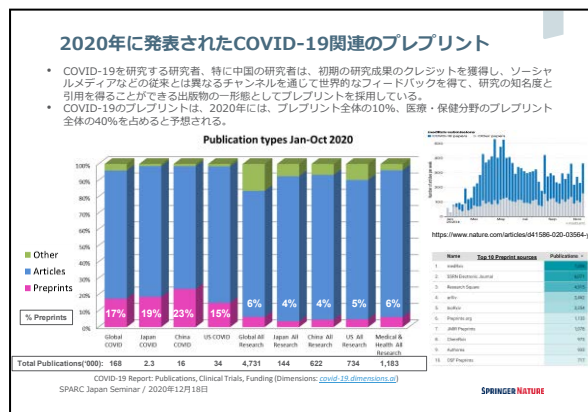
面白いのは、プレプリントがどこで発表されているかということです。Dimensions の中で一番多いのは medRxiv で、あとは SSRN Electronic Journal や Research Square、arXiv です。medRxiv は 2019 年 6 月にスタートしたばかりで、今のところプレプリントの 3 分の 2 が COVID-19 に関するものなので、ほとんど COVID-19 のプレプリントのリポジトリになっています。

プレプリントの長所と短所

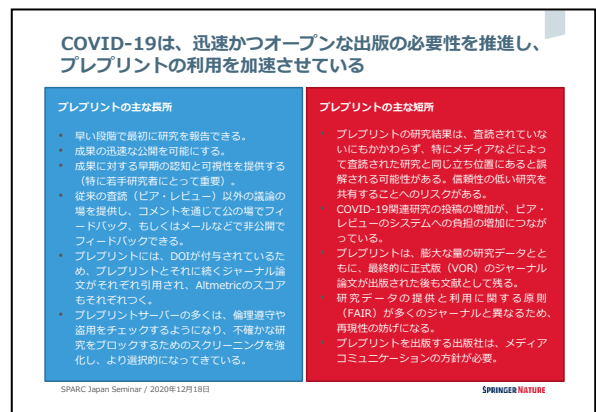
プレプリントの主な長所は、早い段階で「最初がこの研究をしました」という報告ができること、成果の迅速な公開が可能であること、その成果に対する早期の認知と可視性が提供できることです（図 3）。この 3 点は特に若手研究者にとってとても大事なことです。また、従来は、査読のプロセス以外で、投稿する前にフィードバックはほとんど得られませんでした。が、プレプリントは出せば公開もしくは非公開のフィードバックが得られます。また、プレプリントは DOI が付与されているので、プレプリントとそれに続くジャーナル論文の両方が引用され、それぞれに Altmetric のスコアも付きます。また、F1000Research がいい例ですが、プレプリントサーバーの多くが倫理遵守や盗用をチェックするようになっており、不確か



(図 1)



(図 2)



(図 3)

な研究をブロックするためのスクリーニングが強化され、より選択的になっています。

プレプリントの主な弱点は、まず、プレプリントの研究結果は査読されていないにもかかわらず、特にメディアなどによって査読された研究と同じ立場にあると誤解される可能性が高く、信頼性の低い研究を共有することへのリスクが非常に大きいことです。アメリカで結構有名なケースも出ています。また、COVID-19 に関する研究の投稿が大変増加しており、ピアレビューのシステムの負担も増加しています。先にプレプリントが存在するので、科学雑誌は、より早く査読のプロセスを進めて、できるだけ早く査読済みの記事として出したいと考えます。また、プレプリントは膨大な量の研究データとともに、最終的に正式版（Version of Record）のジャーナル論文と全く同じように文献になるので、文献が重複して文存在することになります。それから、プレプリントに掲載されている研究データの提供と利用に関する原則（FAIR 原則）が必要です。F1000Research のようなプレプリントサーバーはきちんとチェックされていますが、全体的にプレプリントサーバーはデータの提供についてはまだ少し弱いのです。それから、プレプリントを出版する出版社はメディアコミュニケーションの方針が必要です。リポジトリも含めてメディアはいろいろなところでプレプリントが見えているので、きちんとしたメディアコミュニケーションの方針が必要ということです。

プレプリント論文のインパクト

プレプリントの数がどんどん増えて、単純に言えることは、インパクトのあるプレプリント論文はインパクトのあるジャーナル論文として出版されるということです。一例を挙げます。10 月の Dimensions データで一番引用されているプレプリントは武漢のグループが 2 月 9 日に medRxiv で出版したのですが、2 カ月半後に同じものが「The New England Journal of Medicine」で論文として出版されています。その引用数は、Dimensions の COVID-19 に関する論文で 2 番目に高い

です。

研究者の目から見て、何がプレプリントの信頼性を高めているかという調査が行われました。「Royal Society Open Science」に掲載された論文によると、欧米の研究者約 4,000 人を調査したところ、プレプリントの信頼性を評価するために大事なものは、基礎となる研究内容の透明さ、例えば研究データへのリンクが透明であること、そしてオープンサイエンスのコンテンツを支持するデータがプレプリントに載せてあることだという結果が出ました。それほど大事ではないのは、著者の所属、著書の名前、グループ、それから、どこのジャーナルに投稿するつもりかということでした。主なプレプリントサーバーには、研究データのリンクなど、透明性を見せるリンクがまだ存在していないこともあるので、プレプリントの信頼性を高めるためにどんどんこのようなスタンダードを insist すべきだと思います。

出版物の撤回

プレプリントについては、さまざまなスキャンダルなどの問題もありました（図 4）。例えばアメリカで二つの有名なケースがあり、結局「Lancet」で出版された論文とそれに関するプレプリントは両方撤回されています。アメリカの大統領がよく言っていたヒドロキシクロロキンに関する有名なプレプリントです。撤回の主な理由は、第三者が研究データにアクセスできないということでした。



(図 4)

しかし、本当に COVID-19 に関する論文の撤回は大きな問題なのでしょうか。今まで 39 報が撤回され、その中の 15 報はプレプリントでした。実は、それは撤回された論文のトータルのたった 2.3%です。ですから、特に大きな問題にならないと思います。撤回は、例えば結果のエラー、データのエラーなど、よく予想できる理由によるもので、プレプリントでも論文の記事でも、COVID-19 に関する論文の撤回は大きな問題ではないようです。

シュプリング・ネイチャーのプレプリントに関する方針

研究データを FAIR の原則でオープンにすることの必要性について先ほど言いましたが、特に COVID-19 の影響で、オープンデータを利用したい人がさらに増えています。シュプリング・ネイチャーの調査によると、他の人のデータを利用したい研究者が大幅に増えているので、プレプリントでもデータのリンクが大変必要になってきています。

2019 年 5 月から、シュプリング・ネイチャーのジャーナルは全タイトルのプレプリントをエンカレッジしています。なぜかという、早い段階で著者のプライオリティや人名が認められるからです。プレプリントはグリーン OA の重要な形態だと思います。査読のプロセスに入る前の段階でできるだけ早く結果が出せるので、出版社にとってはとてもいいことです。

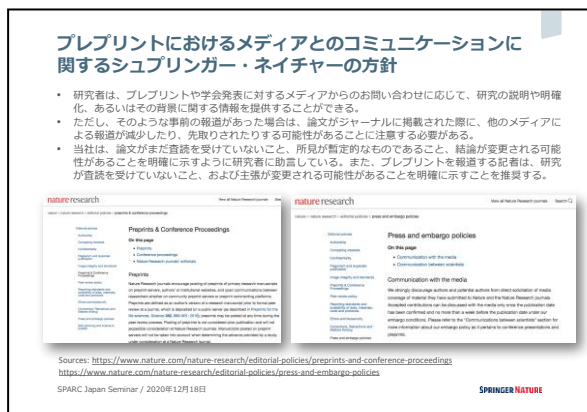
プレプリントにおけるメディアとのコミュニケーション

ョンに関するシュプリング・ネイチャーの方針は図 5 のとおりです。やはり研究者には、プレプリントや学会発表に対してメディアから問い合わせがあれば、必ずそれがまだ査読されていない論文であることを説明していただきたいです。それから、プレプリントの段階でメディアからの注目が大きい場合は、その後、査読された論文がジャーナルに出るところでマスコミの興味が減少するので、もし編集や訂正があっても、正しいデータがそれほど注目されない可能性があります。プレプリントサーバーでも出版社でも、メディアの方針が必要であり、それはとても大事です。

In Review

In Review は、COVID-19 に関する 3 番目のプレプリントサーバーです（図 6）。シュプリング・ネイチャーが出資する Research Square という会社のプレプリントサーバーなので、シュプリング・ネイチャーのジャーナルと In Review はかなり深い関係があります。In Review のプレプリントサーバーの主な特徴は、F1000 と似ていて、プレプリントが存在しながらジャーナルの査読のプロセスを透明にし、査読されている間に早い段階で成果も分かるサービスであることです。

投稿するときには、著者が In Review のサービスにオプトインできます。そして品質管理チェックが行われます。このプレプリントは、著者のバージョンをジャーナルの記事の形に編集して、その段階で公開します。In Review には公開されているプレプリントのバ



(図 5)



(図 6)

ージョンが表示され、ここでは Under Review のバッジが付いています。これは Nature Research の In Review のプレプリントサーバーに載せている記事です。このプレプリントは、雑誌「BMC Developmental Biology」で出版されます。すると雑誌のサイトに記事として移動して、Accepted のマークが付きます。BMC の場合はオープンアクセスの CC-BY のライセンスがあるので、それが In Review のサーバーにも残ります。

最近 Research Square は、In Review の特に COVID-19 の研究によって大変人気になり、2020 年 11 月 16 日で 5 万報のプレプリントを発表しています。過去半年間に発表したプレプリントの 60%は COVID-19 に関するものでした。

データで見る In Review

In Review のプレプリントのデータを見てみると、In Review で出されたプレプリントがそのままジャーナルに受理される割合は 43%です（図 7）。これが高いかどうかは分かりませんが、1 回リジェクトされて違うジャーナルに投稿されるケースも含めると、その割合は大きくなると思います。プレプリントは結構見られており、平均 8 回のダウンロードがあります。最も多くダウンロードされたプレプリントでは約 5,000 回ですが、やはり雑誌の投稿論文よりはダウンロードされていません。

引用については、今までに Research Square で約 2,000 件の総被引用数があります。引用されるプレ

プリントはたくさん引用されますが、ほとんどは引用されていない状態で、全プリプレントのうちたった 3% 未満しか引用されていません。

また、Altmetric のスコアで注目性を見ると、あるプレプリントはたくさん注目されていますが、多くのものには Altmetric スコアがなく、注目されていません。これはジャーナルの記事と似たパターンだと思います。

コロナ時代における査読への負担

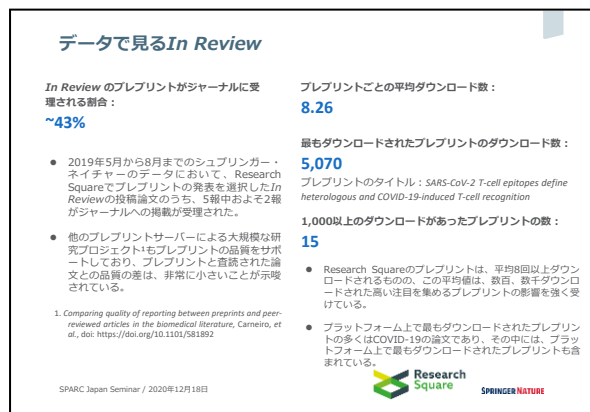
先ほど、ピアレビューのシステムのプレッシャーについて少しお話ししました。COVID-19 の時代で、COVID-19 に関連する論文を出版社はより迅速に優先的に出版したいのですが、質の高い査読は維持したいのです。従って、大変限られた数の査読者に、かなり大きな負担が掛かっていると思います。シュプリング・ネイチャーと他の出版社も入っている C19 Rapid Review Initiative では、ボランティアの査読者のレビューと身元が出版社やジャーナル間で共有できるようになっています。

ピアレビューに関する大変面白いデータとして、COVID-19 に関する論文は大変速く査読のプロセスが行われているのですが、COVID-19 以外の論文はその負担を受けてだいぶ時間がかかったということがあります。

OA ステータス別：2020 年の COVID-19 関連論文

最後に、プレプリント以外のジャーナル論文の OA のステータスを見てみます（図 8）。現在、COVID-19 に関する世界の論文の約 80%は完全に無料アクセスできるようになっています。そのうち純粋なゴールド OA は全体の 20%を占めており、これは世界の全論文のオープンアクセス率である 17%に近い数字です。

なぜ 80%になったのかというと、出版社がクロージドの論文をブロンズ OA の形で無料提供しているからです。それがウィズ COVID-19 の時代で 40%くらいになっていますが、平均的にブロンズ OA は 17%



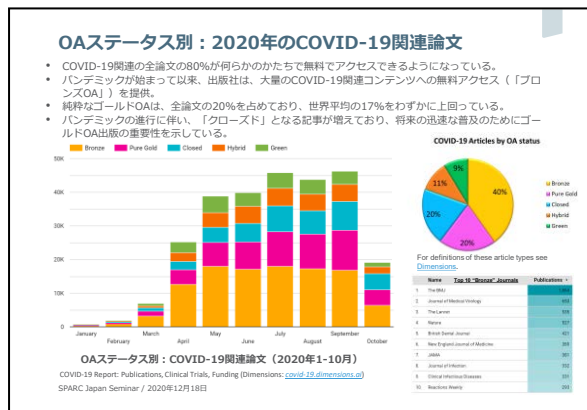
(図 7)

ぐらいです（図 9）。来年、再来年になってアフター COVID-19 の時代になると、ブロンズ OA の影響は減っていくと思います。

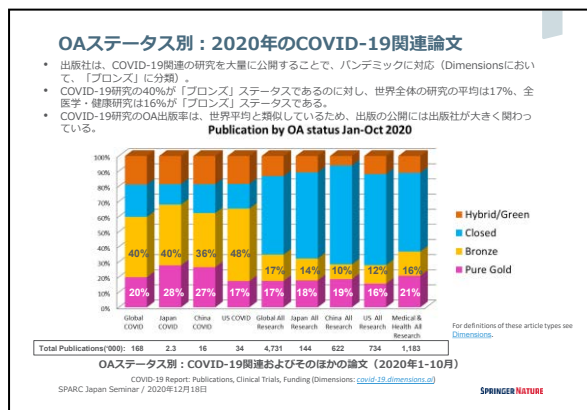
迅速な研究をオープンに公開するためのベストな方法は、ゴールド OA かもしれません（図 10）。COVID-19 に関する研究論文をこれからより迅速に出版し、高いインパクトを達成するためには、ゴールド OA が最適な形態だと思います。

●矢吹 一つ目のご質問は、「図 3 でプレプリントの主な短所として挙げられている、ジャーナル論文が出版された後も文献として残る点について、もう少し詳しくお願いできますか」というものです。

●ブーケ プレプリントサーバーには、ジャーナルへ



(図 8)



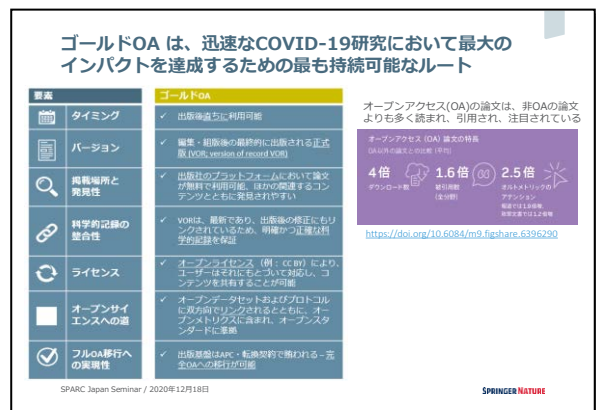
(図 9)

の投稿と査読のプロセスが終わった後もプレプリントとジャーナル論文の両方が文献として残ります。査読のプロセスで記事が改善されたり、新しい意見が出たりするので、若干異なるけれども同じようなものが重複して残ります。たくさん論文の発見（discovery）は必要なのですが、数が増えるので検索する側には大変な負担になるのではないかと考えています。

●矢吹 次の質問は、「シュプリング・ネイチャーではプレプリント論文の引用を支持するということですが、査読前で信頼性が担保されていない状態の内容を引用することへの懸念はないのでしょうか」というものです。

●ブーケ ほとんどのプレプリントは投稿してそのまま雑誌に出版されることになりますが、メディアの方針でよく言われているように、引用する側は「これはプレプリントを引用しています」とはっきり書いていただきたいです。査読のプロセスでこの結論が変わる可能性があることをみんなが理解できるようにするためです。プレプリントの良さは、より早く研究成果などが出ることですが、やはり科学のプロセスでさまざまな改善があるので、読者は引用されているプレプリントだけではなく、最新データもチェックするべきだと思います。

●矢吹 坊農先生から、「図 3 の短所の下から二つ目、



(図 10)

プレプリントに掲載されたうんぬんのところで、FAIR が再現性の妨げになるという書きぶりになっていますが、誤訳ではないですか」というご指摘があります。「FAIR は Findable・Accessible・Interoperable・Reusable にするための原則なので違うのではないのでしょうか。もし元の文章があれば、そちらも示していただけるとありがたいです」とのことです。

●ブーケ これは、プレプリントサーバーは FAIR の原則を利用しないとその再現性が問題になるという意味です。つまり、プレプリントサーバー、リポジトリは FAIR の原則をフォローしなければならないという意味です。

●矢吹 もう一つご質問です。「査読を通らなかった論文、リジェクトされたもののプレプリントは残るのでしょうか。また、不正などによって取り下げられた論文のプレプリントはどのようなのでしょうか」。

●ブーケ その二つのケースはだいぶ違います。リジェクトされたプレプリントは、そのままプレプリントとして残るケースが多いと思います。プレプリントはそのまま文献として存在感があるので、ジャーナルでそのままアクセプトされても、違う形でアクセプトされても、リジェクトされても、そのまま残ります。しかし、不正 (plagiarism, fraud) に関するプレプリントは、撤回のプロセスがあります。また、Retraction Watch はジャーナルの論文だけではなくプレプリントのサーバーも調査しているので、プレプリントも撤回になることがあります。今までは割と少なく 17 報のみですが、その中に結構有名なケースがありました。フランスのプレプリントで、COVID-19 に関するものでした。政府のポリシーに大きな影響があったのですが、それはやはり撤回されました。