

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？」

Preprint が誘導する研究サイクルの力学考

引原 隆士

(京都大学図書館機構長)

講演要旨



arXiv.org が醸成してきた学術コミュニティにおける内的手段であった preprint が、既にこの手段の準備があったバイオ系を含む分野において、コロナウイルス感染下の情報共有及び日々変化する状況への議論の手法として脚光を浴びた。この preprint を含む動きが学術情報流通の新しい手段、もしくは別の手段を生み出すかどうかは、研究のサイクルを生み出す動力学を理解することで、議論することができる。多くの OA ムーブメントの位置づけも合わせて、成熟したからこそ考えるべき、日本の今後の研究の多様性と展開を生む必要条件を論じる。

引原 隆士

1997年以来京都大学大学院工学研究科に所属し、現在同電気工学専攻教授。研究分野は非線形力学の工学的応用、計測とシステム制御、パワープロセッシング。2012年以来京都大学図書館機構長として長きに亘り、オープンサイエンスに関わる活動に従事し、2015年にオープンアクセスポリシー、2020年にオープンデータポリシーを機関として取りまとめた。これらの活動を通じて、オープンサイエンスと同時に図書館だけでなく科学技術活動全般のデジタルトランスフォーメーションを指向している。2016-2018年 arXiv.org MAB。

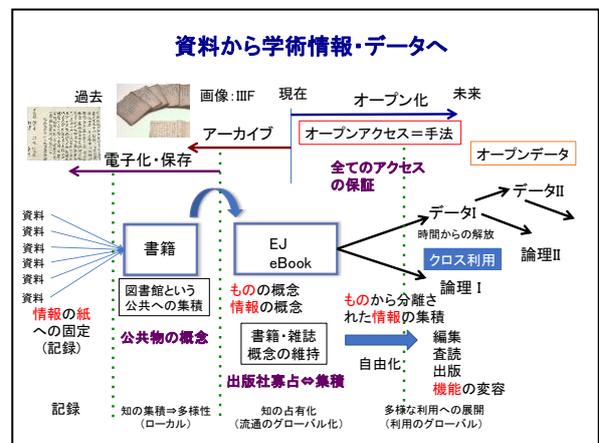


プレプリントは、ステークホルダー間の単純なつばぜり合いではありません。研究活動の活力になるといふ本質的な面を捉える必要があります。そのことを図書館関係者あるいは研究者の方々に理解していただくことが本日の一番大きな目的かもしれません。私は専門分野が数理物理なので arXiv.org をよく使いますが、それに関する話は 2017 年 10 月の SPARC の講演で既にさせていただいたので、今日お話しするストーリーから外させていただきます。

研究のライフサイクルにおけるプレプリント

モデルなくお話しすることは非常に難しいので、まず議論の前提として、学術情報とは何かということからお話しします (図 1)。従来から、情報と呼ばれるものが書籍や論文などに固定化され、ものとして流通

してきました。ものとしてあるということはエビデンスであったわけですが、現在では、大きく見てエビデンスとして論理とデータを考えないといけないというのが学術情報における重要なポイントだと私は認識しています。そのうち、論理とデータに関しては、今は



(図 1)

情報として扱われます。ところが、論理とデータは、ものがなくても成立するので、ものの中に固定化されなくなってきたというのが現状です。

プレプリントを研究のライフサイクルで見えます(図 2)。モチベーションから始まり、研究、下書き、論文と進んでいく中にプレプリントがあり、研究のライフサイクルにおいてプレプリントはあまり大きな意味を持たないとされていると思います。特に図書館関係の方々は、数年前からプレプリントを気にしているかもしれませんが、arXiv.org に関してはもう 20 年以上運用されています。

このライフサイクルの個々のポイントで研究者のどのような能力や技術が必要かということについては、後でゆっくり見ていただきます。研究者にとって、研究を進めてキュレーションをする能力、一般化する能力、さらに論文への落とし込み能力は必須条件です。それに加えてさまざまな技術がなければ、このライフサイクルは回せません。従って、議論になっている論文の部分、サイクルでいう下の方のコミュニティの部分だけを取り上げてプレプリントを考えるのは非常に拙速だと思っています。

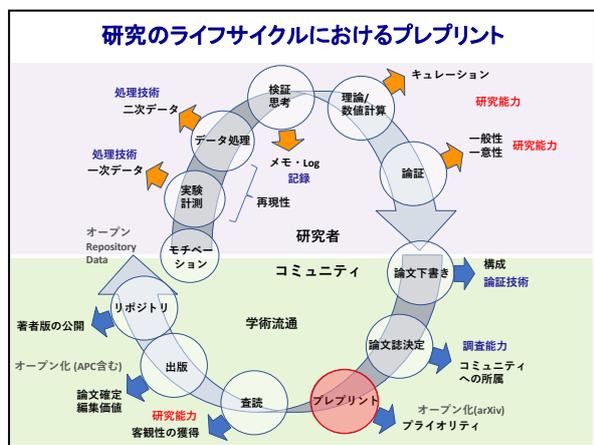
現状は、図 3 のように、プレプリントの範囲が広がっています。本来、プレプリントは小さなコミュニティの中で回す情報流通でした。それが、下書きの段階から論文誌決定、査読に至るところまでもコミュニティで回すようになってきています。先ほど、リポジトリをプレプリントサーバーとする話もありましたが、

そうなるとこの研究サイクルは非常に混乱を来します。ResearchGate も同じで、リポジトリやプレプリントが両方載っている状態の中でこのサイクルをどう回したいのかということが、研究者から見れば非常に重要なポイントです。従って、これをきちんとひもとかないといけません。

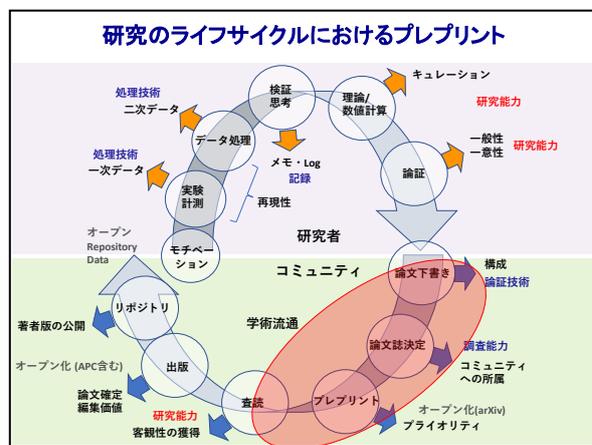
研究・研究者の動向分析

研究者の認識からすれば、研究のオリジナリティは非常に重要なポイントです。オリジナリティを高めるときにコミュニティがあるかないかということは別として、研究を創成し、それを展開して成熟させていく過程の中で、インパクトファクターが礼賛されるフェーズは、人が行ってきた研究をさらに進めていくフェーズです。しかし、オリジナリティを尊重する研究者から見れば、これは終わったフェーズであるわけです。ここで議論していても、研究は広がりはしますが、プラス α 、プラス β と数を増やしていくだけの世界なので、数が増えたからといって、その人を集めてきて研究者として充実しているかという何もできません。どこかでコピーを探してくる人たちのフェーズなわけです。はっきり言って、インパクトファクター礼賛フェーズは、研究者あるいは研究機関から見ると何の意味もありません。ですから、「今までの人事は失敗だった」と多くの大学が言っています。

プレプリントは、創成から展開のフェーズに関与しているということを理解していただかないといけない



(図 2)



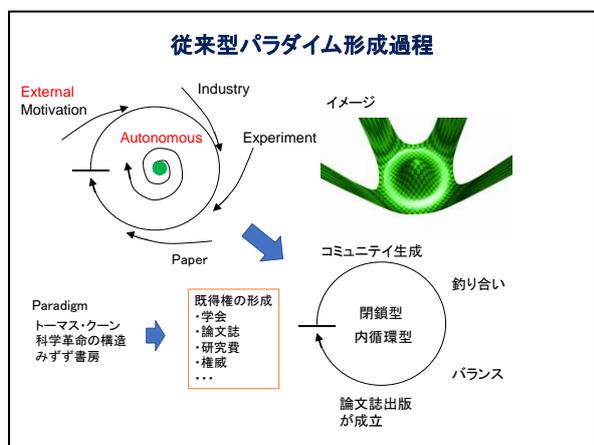
(図 3)

と思います。成熟期間ですぐ出てくるような論文をプレプリントで出しても意味がありません。出版社が「もうすぐ載りますよ。プレプリントも置きましょう」と言うのは成熟期間です。展開期間や創成期間は、その分野に人がいなくて、評価者もいないので、今までの雑誌であれば排除されていた世界です。これをきちんと残していく世界がプレプリントの中にあるのだということを理解していただかないと、何の意味もないのではないかという意見も出てきています。

また、論文がたくさん出てきたからといって、パラダイムシフトが起こるわけではありません。国の政策も大きく間違っていて、はやっている分野に人を集めればいつの間にかパラダイムシフトが起きて新しい製品が出てきて、今まで見えなかったものが見えるようになるという議論がいろいろなところでされていますが、そのようなことは一切ありません。チャージアップしたからといって分野は生まれないということです。商業的には、雑誌の方々、創成・展開の部分は人がいなくて買ってもらえない以上、いくらやっても仕方がないのです。ですからこの分野は、プレプリントが見るべき分野であると言えます。

従来型パラダイムと閉鎖型コミュニティ

少し別の観点から、従来型の学会やパラダイムの形成過程を見てください（図 4）。自発的に生み出した部分が広がっていく流れと、外から入ってくるさまざまな産業的要求、実験結果、他の論文、モチベーシ

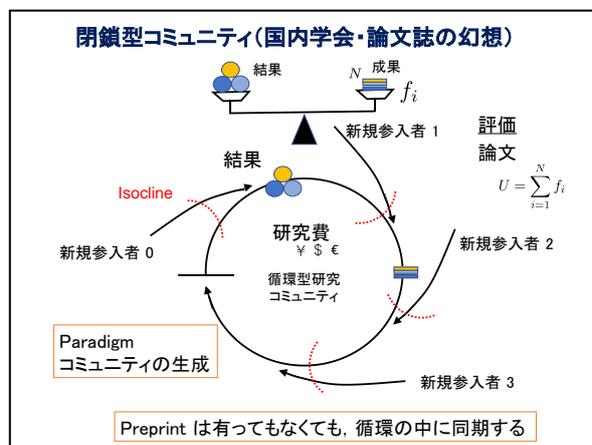


(図 4)

ョンから、研究のサイクルがバランスを取ってぐるぐる回り出し、収束していく過程が一つのパラダイムになっています。イメージとしては、真ん中が持ち上がっていて、あるいは少し下がっていて、外から入り込むようなポテンシャルのたわみがあり、そこでぐるぐる回って自己充足型になっています。Thomas Kuhn のパラダイムの話から言えば、ここで出来上がるものは既得権です。要するに、学会であり、論文誌であり、研究費であり、権威です。これを守ることがパラダイムなのです。従って、パラダイムを守っているのが論文誌であることも事実です。閉鎖型・内循環型コミュニティの形成過程で生まれたものですから、今ある雑誌や学会は、それを維持するためのものであるということを知っていただく必要があります。

たくさんの新規参入者も全て循環型に入ってくるだろうというのが、閉鎖型コミュニティの幻想です（図 5）。研究費はここにあるからみんな寄ってくるだろう、それだけの数の成果をバランスを取って出せますよという世界をつくり上げているわけです。外から入ってくる人も、このサイクルの中でフェーズを合わせながら生活すると研究費にありつけるというのが今までのパターンです。これは私見かもしれませんが、周りで起きているのはこういうことです。

この中では、プレプリントもあってもなくても一緒です。要するに、このフェーズに合わせて一緒に回っていれば、プレプリントがなくても周りに認められていくという状況があります。プレプリントには意味が



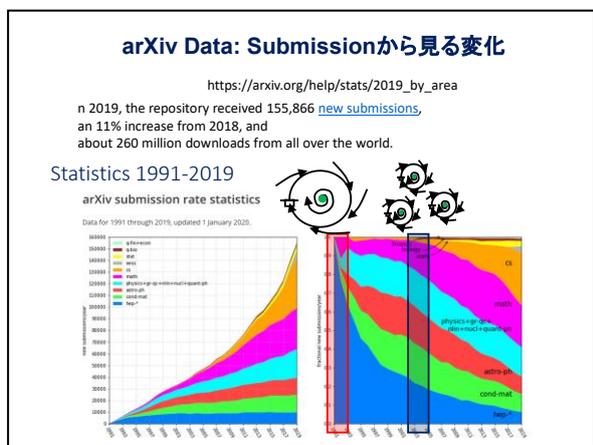
(図 5)

ないと発言される方々は、結局はこの循環の中でぐるぐる回っている人だと言えると思います。

arXiv のサブミッションから見る変化

さて、arXiv のサブミッションのデータから何を見るかということです (図 6)。arXiv の創立の過程から言えば、高エネルギー物理ですから、一つの高エネルギー物理の分野でフェーズが回っていて、その中でコミュニティが形成されます。それがその後、いろいろな分野に分かれてくるわけです。もちろん数が変わっているので、比率だけで見るとあまり正確ではないですが、この分野の中でそれぞれ別のコミュニティの中で回り出すものを、真ん中の辺りはサポートしていたと言えると思います。

しかし、今、この内循環型のパラダイムが変わりつつあります (図 7)。変わりつつあるというか、既に変わっているわけですが、要するに世界中で学会等がタコつぼ化する、あるいは学会が学会誌を出版社に買い取られて、そのコミュニティが小さくなっている、コミュニティが存在しなくなっているということが起きています。結果として、その中に若い人が入ってこなくて研究者が老化する、あるいは予算が減って既得権が小さくなり人が小さくなっていく。そうすると当然、中の引力は小さくなりますから不安定化します。そして外からいろいろな情報が入ってくると、中にいる人たちは外のものを求めて出ていくという状況が起きてきます。コミュニティが維持できなくなって、外



(図 6)

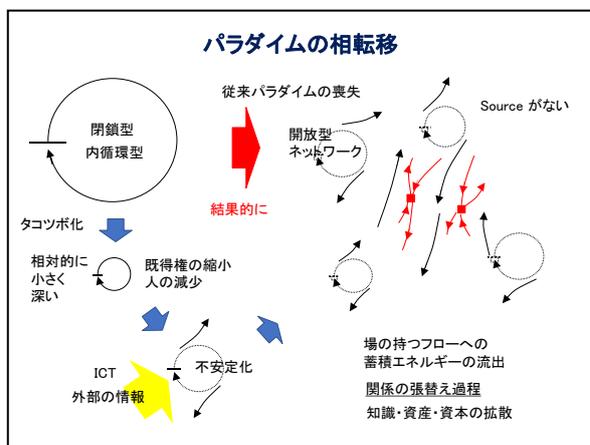
に人を発散し、予算も発散していく、あるいは論文誌も総合的なものも変わっていく。その中では、境界はできるかもしれませんが、フローが起きてくる。それが代替的に力学的に見た流れになるかと思います。同じようなことが arXiv でも起きています。

統計を見ると、サブミッションはどんどん増えているのですが、それ以上にダウンロードがものすごい数になっています。これは新しいものを求めてきていることを示しているのですが、一つの分野の人だけではなく、何かいい解析方法はないか、いいデータはないか、いいアルゴリズムがないかということを探められるのが arXiv のデータベースとしての在り方です。

要するにここで起きているのは、コンピューターサイエンス、math など、横断的なものが広がってきて、高エネルギー物理や凝縮系の部分の比率が減り、どちらかという全体としてフローを起こすような研究内容が広がっているということです。それが今求められている arXiv のプレプリントとしての在り方です。

開放型コミュニティの出現

それを絵で描き直すと、図 8 のようになります。arXiv は一つのパラダイムを作っていました、それが開放型のコミュニティになってきています。そういうところは元々あったのですが、オリジナルな方々もプレプリントを持って外に出ていき、それを他の研究と合わせていくようなことが起きています。新規参入の方々も、一時期ここで in phase するのですが、プレ



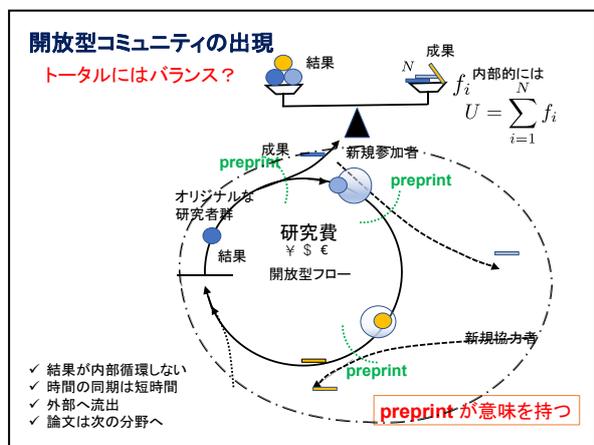
(図 7)

プリントという成果を持って出ていくという形です。そして協力者も同じような形で、あるときだけプレプリントを書いて、外に出ていく。つまりプレプリントは、ここにいる一瞬の間だけの in phase を保つための手段になっているということです。

固定的な流れの中ではプレプリントとジャーナルはほとんど変わらなかったのですが、分野的には、理論物理系などは時間もかかりますし、ジャーナルを待っている暇はありません。そのコミュニティの中だけなら皆さん一緒に回るので何も問題ないのですが、少し寄ってまた外に出ていく人には待てません。だからプレプリントで成果を残して外へ出ていく。プレプリントはそういう大きな意味を持っているということになります。結果として、内部循環がもう存在していませんから、ある一瞬の時間、その切り出しが大変重要になります。論文は次の分野へ行ってまた書くということにもなり得るわけです。

受け皿の乱立

アーカイブ、プレプリントサーバーが山のようにあります。開放型になった結果として受け皿が乱立しているのですが、この中で権威があるかないかという話も出てきます。これは「乱立」でも、「受け皿がたくさんある」でも、どちらの言い方でもいいかもしれませんが、いずれもが閉鎖型や内部循環型のシステムにはなっていません。bioRxiv でも数理的なものがたくさんありますし、最近の COVID-19 でも数理的な感染



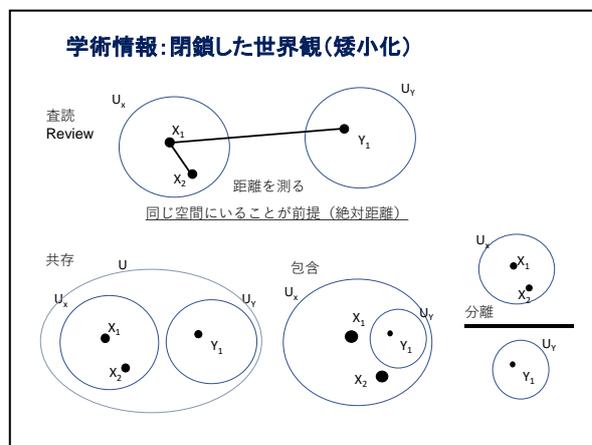
(図 8)

拡大のモデルなどの話は bioRxiv にも出ますし、arXiv にも出ますし、他のところにも出ています。それらをどうやってリンクするかという話は重要ですが、どこでもいいわけです。要するに自分たちのメインのところ寄って、そこで「これはこんな効果がありますよ」と示すことが重要なので、そこから出て論文誌だけ別のところで出すことは当然可能です。ですから、この世界自体がもうフローの中にあると思います。

言いたいことは、それぞれの受け皿が固定的に権威を持って残るかどうかというのは別のことだと思いますが、今この情報流通が非常に必要な時代になっているということです。きつい言い方をすれば、出版社の方々はこの何を何とか取り込もうと網を張っています。しかし、そうではなくて、このフローを維持しなければ、科学の世界はまた既得権の収奪の世界になってしまいます。フローを維持することを考えなければ研究は進まないと思います。

学術情報：閉鎖した世界観・開放された世界観

学術情報が閉鎖した世界観を描くと、図9のようになります。論文を出した人たち (X_1 , X_2 など) があって、別の人 (Y_1) も出したとします。そのときに、同じ空間にいるかどうかの距離を測るのが査読です。同じものである必要はなく、自分たちと同じパラダイムの中で動いているかどうかを見るのが査読だと思います。その結果、「同じ世界にあるから共存しましょう」「一緒だから包含しましょう」となる、あるいは



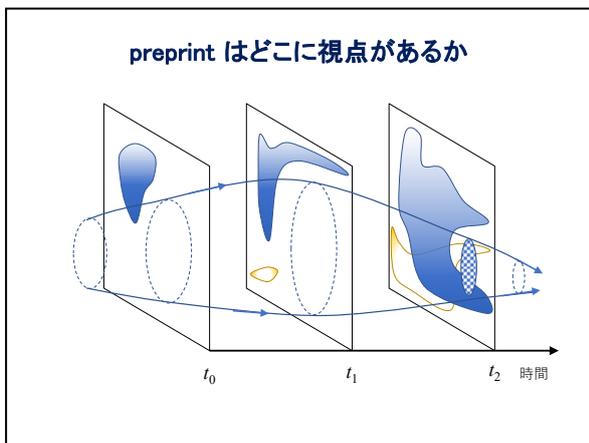
(図 9)

「違うから別の世界に行きなさい」と切り出してしまいうわけです。これが今までの閉鎖した学術情報流通の世界観です。論文誌というものが、そのレッテルになるわけです。

ところが、グローバル化されたら、そうではなくて、ある一定の距離を保ちながら相互の関係を満たさないといけなくなります。そこで相互の関係を測るメジャーが必要になります。このメジャーがプレプリントです。あるいは論文でもいいのかもしれませんが、それぞれ同じ世界にいるわけではないので、共通で見えるメジャーが必要になります。それが今起きていることで、別の世界へ行っても自分が測ってもらえるようにメジャーのポイントとなるものを残していく。その行為がプレプリントになっていると私は研究者として見えています。

プレプリントはどこに視点があるか

話をもう少し進めて、プレプリントはどこに視点があるかということを考えてみます(図 10)。研究が始まり、創始的なものから最後に結論が出ていくという時間経過の中で、オリジナリティがある過程と、広がっていく過程があります。その中で、研究のロングスパンをまず描きます。オリジンからターゲットがだんだん広がり、包括領域が広がって、最後はその応用領域として収束していきます。この中で必ずあるのが、この視点の中に入っていない外側の成果です。それを、自分たちがこの中だけのロングタームを肯定するため



(図 10)

に否定してはいけません。否定しないために、外側にあるものをきちんと残していけないというのが科学の世界、研究者の世界です。

青で描いたものが広がっているのに、自分たちの世界観だけにいると、それが収束していったときに全くずれてしまうということが起きます。プレプリントは外側にある世界を見いだすものだということも意識しないといけません。中のものでプレプリントを置いていくということは、時間としてどちら側にあるかだけの話で、あまり意味がないのです。

研究者的な研究促進とファンドの研究ターン促進

研究者が研究を促進したいときは、研究開始から結論が出るまでの時間を短くします。ですから、人が集まっているところで書き出して、人が去っていく前に終わらせてしまうのが、一番賢い、世の中をうまく回る研究者のパターンだと思います。そして目標を達成して予算をもらって出ていってしまいます。

しかも、この研究者の性(さが)はファンドにうまく利用されています。選択と集中ですが、ほとんど結果が見えてからファンドを与えて、みんなが頑張ってくれる間に成果を評価します。別にファンドがこの分野を育てているわけでも何でもなく、既に育てているのです。育てるのだったら、誰も見てくれないところや、始まる手前のものを本当は育てないといけません。それにはファンドは何の効果もありません。ですから、ファンドがプレプリントを要求したところで、そんなものは意味がないと私は思っています。プレプリントの外側にあるものが重要であり、プレプリントは、目の目を見ていないところにきちんと残していくことが重要です。なぜかということをお話します。

図書館的視点

図書館が現在プレプリントとして意識しているものは、既存パラダイムの外にあると思います(図 11)。

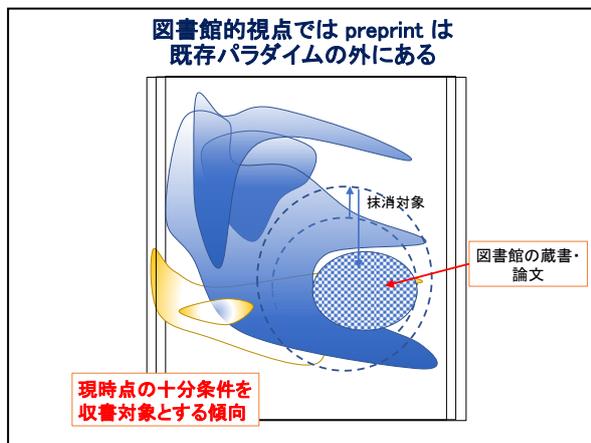
図書館の蔵書や論文は、非常に狭いもので、最終形として残ったものです。それはパラダイムとしてぐるぐる回って成熟したのですが、これしか残りません。ところが、科学の世界観、グローバリズムからいえば、その外側に重層的にさまざまな成果が上がっているわけです。それを否定してこれを残すのかというのが、今までの論文誌の在り方の問題です。

プレプリントとして、自分の時間を提示しながら、こちらが手前で向こうが後なのだということもきちんと残しながら置いていくことが、科学の透明性として非常に重要です。成果が出たものに対してものを言うことはいくらでもできますが、それは後付けの理屈です。本当は、この過程が広がって狭まる、外側ではないということを示した点線の部分を残していくことが科学の成立なわけです。ところが、現時点での十分条件だけを求めているのは、論文誌も出版社も変わりません。それが現実で、皆さんがされているのは営利事業だと思います。

arXiv が示す原点

arXiv が示した原点は、要するに研究をエンカレッジするのだということです。ですから、ステークホルダーが取り合いをしているという話ではなく、各研究者が研究できるようにするという事なのです。アーカイブやプレプリントの議論の中では、その観点が抜けていると思います。

もう一つ重要なことは、クリアでシンプルなインタ

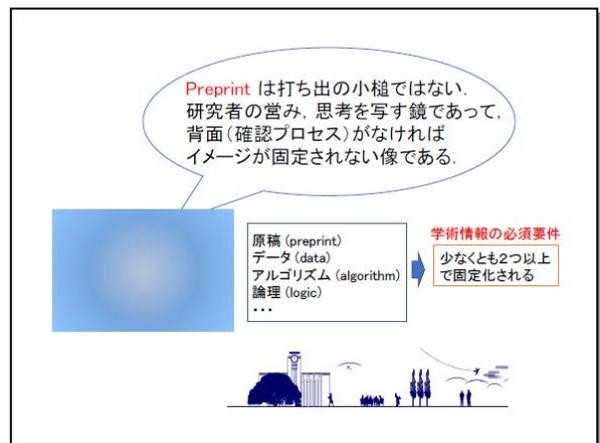


(図 11)

ーフェースであるべきだということです。いろいろなものが乱立している中で、出版社自身が「プレプリントを置いたら載りますよ」と言うこと自体が大きな間違いだと思います。それは研究者の活動の取り込みにすぎません。営利活動だからそうなるのかもしれませんが、それは研究活動を支援するものではないと思います。研究者は、研究分野のアーカイバーではありません。研究者は過去のものやずっと考えていく者ではなく、まだ見ぬ世界を探検する者ですから、それをエンカレッジするようなものでなければならぬと思います。そして、プレプリントがその可能性を持っているものであり、その外側の世界が本当は探求する世界だということを皆さんはあまり意識されていないということ、あえて申し上げました。

プレプリントは研究者の営みや思考を映す鏡

当然ながら、プレプリントは打ち出の小槌ではありません。研究者の営みや思考を映す鏡であって、背面（プロセス）がなければイメージが固定されないものだと思います（図 12）。つまり、今見えるプレプリントの評価はジャーナルに射影したものであって、ジャーナルの観点からプレプリントが見られています。これは間違っていると思っています。ジャーナル自体が研究の一断面にすぎず、プレプリントはもっと広い世界を持っているのだということを見たら、むしろプレプリントの射影がジャーナルにすぎないのではないかというのが私の意見です。



(図 12)

また、学術情報の必須条件は、原稿やデータ、アルゴリズム、論理などのうち、少なくとも二つ以上で固定・検証されるものです。一つだけで検証されるものは科学としては成立しません。それはプレプリントでも変わらないということを最後に申し上げたいと思います。

ので、ご質問はすぐには出づらいのかもしれませんが、大変勉強になりました。

●引原 プレプリントを現象論で見るのか、手法論で見るのか、観念で見るのかで大きく違うと思うので、今役に立たないというような判断をするのは、関係者はやめた方がいいだろうというのが私の基本的な姿勢です。

●矢吹 深貝先生から、「引原先生の創成段階に焦点を当てた強調は、社会科学の側からも見て分かりやすいものです。力学的観点でご説明いただきましたが、知の進化論といった設定も可能です。さて、これは後のパネルに向けてですが、コロナ状況という新規事象の中で、既存の細分化を越えた領域横断を迫るという観点でプレプリントを位置付けることができれば」という、この後のパネルに向けてのコメントが来ていますが、もし応答があればお願いします。

●引原 おっしゃるとおりだと思っています。プレプリントは、大きな変化をするときに非常に強力な武器になると思います。COVID-19 がそれを示しているように、以前ですと、高温超伝導が発見されたとき、ほとんど一晩でいろいろな成果が変わっていくときに、社会も大きく変わる意見もたくさん含みながらプレプリントが発散的に広がった時代があります。同じように、大きく変わるときであればこそプレプリントが意味を持ち、論文誌が役割を失うという現実があるということを見たときに、社会が定常で安定した状態であれば、多分ほとんどなくなると思います。しかし、社会が発展するためには安定を続けてはいけなくて、安定が一回不安定になって次の安定を求めていくという発展形を作らないといけないとすると、プレプリントのある世界を広げて、そこで見ない限りは伸びないだろうというのが私の個人的な意見です。

●矢吹 ダイナミズムを整理していただいたところな